The Dissertation Committee for Jennifer Fang En Lee Certifies that this is the approved version of the following dissertation:

PROBING APTAMER SPECIFICITY FOR DIAGNOSTICS

Committee:

Andrew D. Ellington, Supervisor

Jon Huibregtse

Lauren Meyers

Philip Tucker

James Bull

PROBING APTAMER SPECIFICITY FOR DIAGNOSTICS

by

Jennifer Fang En Lee, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2007

Dedication

To my family, without your support and unconditional love I would have never made it this far. You are the anchor of my life that gives me strength in all of my endeavors.

Acknowledgements

I want to first thank Dr. Andrew Ellington for providing me such an opportunity to learn and grow. Through his teachings I have not only gain valuable knowledge about science but also how to grow and mature as a person. I also want to thank my committee members, Dr. Lauren Meyer for her generosity at providing me the first insight into the field of theoretical biology, for allowing me to work with her the at the beginning, and for always being such a wonderful role model. Dr. Jon Huibregtse, who is never too busy to speak with me and educate me more about cancer biology either through meeting with him or through his Tumor Biology class and class notes, a field that I will continue pursuing many more years to come. Dr. James Bull who has always provided me an alternative view of my data and never cease to provide his support and show his sense of humor on whatever the data may bring. Dr. Phil Tucker, another great role model who always put the students first, and represents someone who's not only a leader in this field and still balance a great lifestyle of fitness, something worth emulating.

My achievements in experimental science would have not been possible if it weren't for the members of the lab. Everyone in the lab has played a pivotal role in my laboratory experience. In particular, Michael Roberston, who patiently introduced me to the basics of experimental research from pipetting to conducting my very own research. Ted Chu, who has unselfishly shared his trade secrets in cell biology and introduced to a

whole new discipline of research, Cancer Biology and also how to have a good time working with the cells.

To the robotics team, Colin Cox, Gwen Stovall, and Travis Bayer who showed me how working with robots can be rewarding and frustrating at the same time; but nevertheless always with the positive attitude that no mechanical failures will occur again. Matt Levy, my bench neighbor, who not only shared his mentorship but also his good sense of humor. Charles Wei and Tim Tsai who are undergraduate students and have shown great work ethics in helping me with many of the automated selections. Carlos Garcia and Anna Szanfranska, the previous robotics facility manager, who has made sure that the facility would run smoothly so that the rest of us could carry out our experiments without glitches. Dat Vo, who would always ensure that I did not miss the latest literature published pertaining to my field of study. Brad Hall, who provided me help with primer designs and bioinformatics discussion.

For those who have kept me sane all these years, Letha Sooter, who has never stopped believing in me and never failed to share words of wisdom in times of hardship. I also want to thank Supriya Pai and Amrita Singh who have never failed to make me smile and to provide me with any support possible, difficult days are definitely not so unbearable when there are friends like them to speak with at the end of the day. Amos, who has never failed to provide help when the occasion called for it. Finally, a big thanks to Tony B. for his support and for his display of courage and strength in times of hardships allowing me to realize that there is no such a thing as a problem that cannot be solved.

PROBING APTAMER SPECIFICITY FOR DIAGNOSTICS

Publication No._____

Jennifer Fang En Lee, Ph.D.

The University of Texas at Austin, 2007

Supervisor: Andrew D. Ellington

Theoretical studies focusing on the nature of landscapes that correlate molecular sequences to molecular function have mainly been carried out *in silico* due to the vast amounts of data that are needed for analysis. *In vitro* selections of aptamers are a good model system to study theoretical questions at a experimental level. With the introduction of robotic platforms that conduct *in vitro* selections, it is now capable of producing significant amounts of data in a short time, making theoretical modeling with real experimental data attainable.

I will be using a Biomek 2000 Laboratory Automation Workstation to carry out multiple in vitro nucleic acid selections in parallel. I will explore the sequence space to examine whether existing *in vitro* selection systems are optimal at isolating the best winning species. New methods will be introduced that will allow for the selection of identical targets with identical pools free of cross contamination on the open robotic system. This will open the doors to further conduct selections against other identical or

highly similar targets, such as complex cellular targets.

Finally, I will investigate the methods to improve the effectiveness at isolating aptamers against the highly complex lung cancer cell lines. These targets are highly challenging for isolating specific aptamers because of the great diversity of biomarkers found among them. Moreover, their highly morphological similarity of the cultured cells makes selections for specific aptamers very difficult. I explore the different methods that will allow for the generation of aptamers that can distinguish between non-small cell lung cancer and small cell lung cancer, and between non-small cell lung cancer and normal lung cells. Fine-tuning of this process is essential at transferring this process to automated platforms for large-scale generation of biosensors against tumor biomarkers.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

BACKGROUND OF APTAMERS

It has been over 15 years since the modern version of *in vitro* selection was codified [1, 2], and during that time, aptamers have been selected against an extremely wide variety of targets ranging from small inorganic molecules to whole organisms [3-6]. In part because aptamers have proven to have high affinities (picomolar to nanomolar dissociation constants) for their cognate targets and specificities that are comparable to those of monoclonal antibodies the development of aptamer therapeutics has now begun to take off. **Table 1.1** summarizes the important differences between aptamers and antibodies and the advantages that each one confers. **Figure 1.1** denotes the immense size difference between an aptamer and an antibody.

Numerous aptamers have been selected against therapeutic targets such as IgE, IFN-g, alpha-thrombin, PTPase, and others, and have shown great efficacy in tissue culture experiments and animal models. In earlier years, an anti-thrombin aptamer was used in place of heparin for anti-coagulation during heart bypass surgery in canines [7], while aptamers against inflammation factor human neutrophil elastase (hNE) were shown to significantly reduce lung inflammation in rats and had better specificity for their target than an anti-elastase IgG control [8]. Several aptamers have just begun to reach the clinic. This number should increase greatly in coming years, in large part because methods for the delivery of nucleic acid therapeutics are now being developed, and because the pharmacokinetic properties of aptamers have begun to be explored in detail.

Table 1. Unique Properties of Aptamers

	Aptamer	*Antibody*
Selection Process	Versatile in the type of target	Target site is immune system dependent and to select against toxins or non-immunogenic targets
	Different environmental and selection conditions can be designed	Limited to physiologic conditions
	High-throughput capabilities	Time consuming
Properties	Kd in the nM range or lower	Binding ranges from pM-nM
	Little variation in binding	Large variation in binding from different antibody preps
	Small and can be mass produced at a reasonable price	Expensive
Application	Wide range of modifications can be applied	Difficult to modify antibodies
	Exceptional stability	Limited shelf life and highly sensitive to temperature fluctuations
	No evidence of immunogenicity	Significant immunogenicity
	Existence of an antidote to reverse the effect of the aptamer	No antidote available

Table 1.1. The table depicts the differences between aptamers and antibodies from their selection process to their applications.

THE SELECTION PROCESS

An initial pool of about $10^{14\text{-}15}$ different nucleic acids species is incubated with the target of interest. After an arbitrary incubation time, the unbound nucleic acid species are partitioned away from those bound to the target. The bound aptamers are eluted from the target and further amplified through a series of amplification steps producing an enriched pool to be carried into the next round of selection. This iterative process is

repeated until an increase in pool binding is observed. The process of selection is illustrated in **Figure 1.2.**

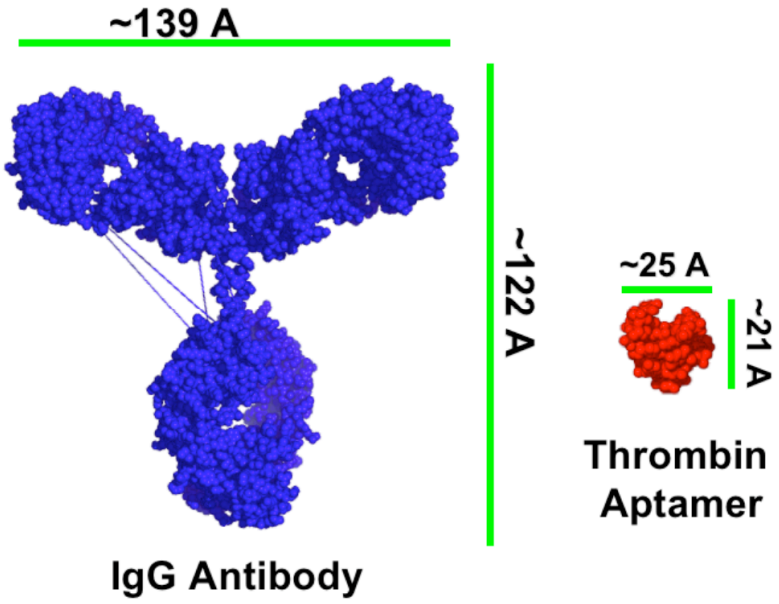


Figure 1.1. Antibody / Aptamer Size Comparison. An estimated comparison of the size difference between an antibody (human IgG) and a selected aptamer (anti-thrombin DNA aptamer) is shown with space-filling models of both. It should be noted that the anti-thrombin aptamer is only 15 residues in length [9], while the anti-gp120 aptamer that has been shown to bind sterically hidden epitopes is 77 residues in length

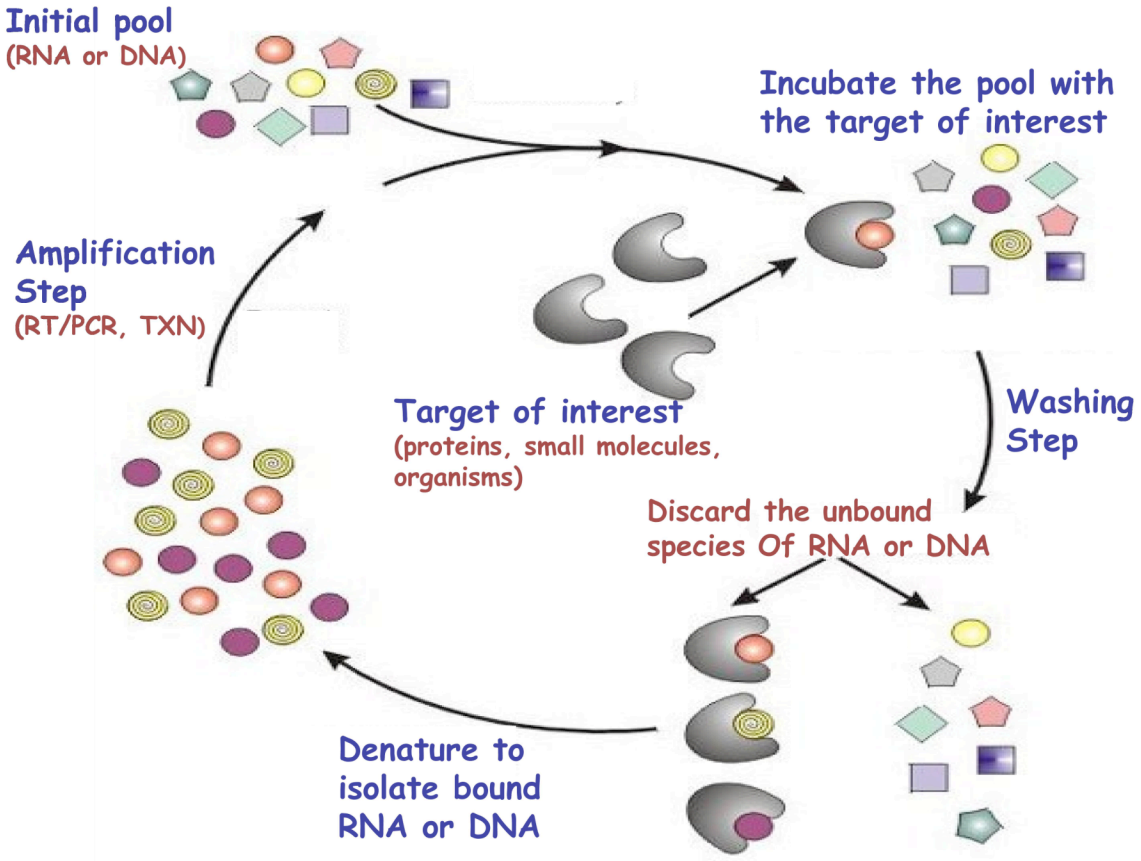


Figure 1.1. Selection Schema. A random RNA pool containing ~10^{13} – 10^{15} unique individual sequences is incubated with a target. The pool / target complex is washed in order to remove low-affinity nucleic acid. Next, high-affinity RNA is eluted from the target and reversed transcribed, PCR amplified and transcribed. This lower diversity pool is then incubated with a new aliquot of the selection target, and the selection cycle begins again.

APTAMERS IN THEORETICAL STUDIES

RNA aptamers targeted to small molecules have served as model systems for the study of evolutionary theories of macromolecular interactions. Aptamers are great candidates for theoretical studies because they can be subjected to mutational pressures that affect one single property such as binding. The Lehman group has shown that *in vitro*

selection can be used to explore the topography of fitness landscapes for the sequence space surrounding a given RNA, either for its original activity or for acquisition of a new activity [10]. They evolved *in vitro* a group of ribozymes and found that these molecules had smooth phenotypic landscape, in the presence of a rugged genotypic landscape [10]. Structural features that contribute to ligand binding make RNA aptamers the choice molecule for these studies. The Famulok group selected aptamers to the amino acid citrulline, and then evolved these aptamers to bind to a related molecule arginine [11]. Through only three mutations, aptamers to arginine were isolated, and these aptamers had secondary structures similar to citrulline aptamers [11]. These mutations not only introduced codons for arginine but were later shown by NMR to make close contact with the target [12]. Dopamine-binding RNAs were evolved to recognize another related amino acid tyrosine, however, in contrast to the arginine/citrulline aptamers, more mutations were needed to change the specificity from dopamine to tyrosine leading to unrelated RNA structures [13]. These studies focused primarily on the end products of the selection, without addressing evolutionary intermediates. Local accessibility of new functions between related sequences of aptamers can be revealed through a systematic examination of a defined sequence space.

PROBING SEQUENCE SPACE

Theoretical and experimental analyses have offered complementary perspectives on neutral evolutionary models by demonstrating that sequences differing by only one or two positions (Hamming distance = 1 or 2) can be structurally equivalent, thereby forming neutral networks in sequence space [14-16]. Neutral network in structural

sequence space referred to RNAs of a fixed length that possessed similar or related secondary structures [16]. These secondary structures served as the bridge for unrelated regions of sequence. To travel through sequence space, sequences will travel through a structural neutral path by the accumulation of mutations. A single sequence capable of two distinct functions is said to represent the intersection between two neutral networks. The Bartel group found the existence of an intersection sequence by evolving a ribozyme with dual catalytic functions [17]. This ribozyme represented the neutral path that led to two evolutionary distinct ribozymes. The RNAs' ability to adopt multiple conformations makes intersection sequences highly probable when the secondary structure is the phenotypic criterion.

Held et al. provided first detailed examination of the complete intervening sequence space between related but functionally distinct aptamer RNAs [18]. They showed the conversion of flavin-binding aptamers to a diverse collection of GMP-binding sequences through *in vitro* selections. They probed the fitness landscapes between two FAD aptamers and between an FAD aptamer and a GMP aptamer by assessing ligand binding by all evolutionary intermediates. Their results suggest that RNA sequence space contains many unrelated solutions to the problem of GMP binding and that some GMP aptamers are very near in sequence space to neutral networks for FAD-binding function. When they carried out structural analysis of closely related FAD-binding and GMP-binding aptamers they observed very few structural differences until the final switch in target recognition was attained. These findings show how an FAD-binding phenotype could be maintained in a rugged landscape through mutational drift along a neutral network until the critical "off-path" mutations would force the transition

in the structure to the nearby GMP-binding phenotype. All these findings further support the neutral evolutionary theory.

Since aptamers can serve as ideal candidates for theoretical studies such as probing sequence space, generation of vast amounts of data would be ideal for the task. Generating enough data for this type of study might have been considered impossible, however with the advent of automated selection [19, 20], achieving this goal is not so far fetched with slight improvements and careful planning.

CELL SURFACE SELECTIONS

Reagents that can recognize vast amount of cell surface biomarkers of tumorigenic cells are crucial in the early detection and improved prognosis of the disease. Cell surface signal transduction proteins such as receptor tyrosine kinases (RTKs) play a key role in regulating cell growth and proliferation in several different types of cancers [21], making them attractive targets for ameliorating tumor progression. Mutations in the RTK known as RET are responsible for the onset of both Type 2A and 2B Multiple Endocrine Neoplasia (MEN) syndrome and familial medullary thyroid carcinoma [22, 23]. Nuclease-resistant RNA aptamers specific to mutant RET were obtained by carrying out selections against RET-expressing cells [24]. These aptamers were shown to inhibit the activity of the protein on the surface and to block downstream signaling. Based on these results, it was hypothesized that the aptamer functioned by preventing the formation of the RET dimer. Since numerous RTKs and cell surface receptors function as dimers or oligomers, additional aptamers might also prove to be inhibitory by a similarly elegant mechanism.

Other extracellular proteins involved in signal transduction can be targeted. For example, increased expression of alpha-beta-3 integrin can lead to angiogenesis and is correlated with multiple human tumors such as melanoma [25]. An aptamer selected against the alpha-beta-3 integrin has been shown to inhibit angiogenesis in vasculature by down-regulating integrin-activated signaling pathways [26].

Aptamers can not only be selected against purified targets or antigens, but also against heterogeneous mixtures of targets, such as whole cells (recently reviewed by Yan *et al*. [6]). Selections against complex targets allow aptamers to be generated even when biomarkers are not known in advance, and should facilitate the identification of new biomarkers. For example, aptamers targeting human osteoblasts from an osteosarcoma tissue were selected and used to coat the tissue culture plate wall to directly capture the osteoblasts in solution [27].

Negative selection procedures with whole cells can yield aptamers that finely discriminate against different cell types. For example, the selection against the RET receptor tyrosine kinase described above included negative selections against cells that did not include the RET-expressing vector and against cells that expressed a mutant form of RET [24]. In one of the most intriguing examples to date, Homann and Goringer [28] were able to select aptamers against whole trypanosomes; and, the target was localized to the flagellar pocket by photo-crosslinking experiments. Later selection experiments that specifically targeted VSG also yielded aptamers that could bind tightly and specifically to the surface of the parasite [29]. Complex targets are not restricted to only cells or organisms; selections have been carried out that targeted plasma and aptamers were coordinately isolated against 14 different plasma proteins [30].

Homann and Goringer [28] selected aptamers against whole trypanosomes; these were eventually found to bind to the variant surface glycoprotein. The target was initially identified by photo-crosslinking experiments, which allowed the determination of its size; fluorescently labeled aptamers could be localized to the flagellar pocket of the organism. It was suspected that the protein target was in the variant surface glycoprotein (VSG), and later selection experiments that specifically targeted VSG also yielded aptamers that could bind tightly and specifically to the surface of the parasite [29]. Similarly, Pan et al. [31] generated both RNA and modified nuclease-resistant RNA aptamers to Rous sarcoma virus (RSV). These aptamers were able to significantly reduce the rate of RSV infection in a quail fibrosarcoma cell line at a concentration of only 20 nM in serum.

The ability to select aptamers against simple or complex targets and their seemingly pluripotent binding abilities make them excellent reagents for the study, diagnosis, and perhaps even therapy of cancers, as reviewed by Cerchia et al. [32]. In fact, selection experiments have yielded RNA and DNA aptamers that can bind to cell surface targets on tumors [33-35]. In each instance, the aptamers were able to selectively bind transformed cells but not normal cells. For example, Hicke et al.,[34] targeted human U251 glioblastoma cells with a 2'-fluoropyrimidine, modified RNA pool. Selected, modified RNA aptamers were found to bind the extracellular matrix protein tenascin-C (TN-C), a protein that is believed to be a hallmark for the onset and metastasis of cancer. The selected aptamers formed complexes with TN-C and had a dissociation constant of 5 nM. The aptamers also bound tumor tissue expressing tenascin-C, but did not bind tissue that lacked TN-C. In fact, the aptamer was also able to discriminate (by 20-fold) against mouse TN-C, even though this protein shares 98% sequence identity

with the human protein. Other selection experiments have also revealed that aptamers are frequently highly species specific (see, for example, White et al., 2001 [36]).

In a second example, Lupold et al. [35] targeted prostate-specific membrane antigen (PSMA), a membrane-bound glycoprotein that is found in prostatic epithelial cells and that is overexpressed by prostate cancers. In fact, its detection in increased quantities is considered a hallmark of the disease. PSMA has been shown to have folate hydrolase activity as well as peptidase activity although its actual biological function is not known [37]. Two aptamers were selected from a 2'-fluoropyrimidine, modified RNA library that spanned 40 random sequence positions. The aptamers were distinct from one another and likely bound different epitopes of the PSMA antigen. The aptamers could inhibit the peptidase activity of PSMA with K_i values in the 2-10 nM range. Fluorescently labeled aptamers were also shown to bind to a prostate tumor line (LNCaP) that normally overexpresses PSMA, but not to PC-3 cells, a different prostate tumor line that does not express the antigen.

Finally, Blank et al. [33] carried out a selection against rat endothelial cells immortalized with adenovirus, and isolated a single-stranded DNA aptamer that proved to be specific for glioblastoma. The aptamer was found to bind to the protein pigpen, which plays a role in angiogenesis by regulating endothelial cell proliferation. Pigpen expression is up-regulated in actively dividing cells and is down-regulated as they become confluent, suggesting that pigpen helps regulate endothelial cell differentiation [38]. Pigpen has both a transcription activation domain as well as RNA binding motifs which are thought to interact with positive cell cycle regulators. Given the way in which the selection was carried out, it therefore appears as though pigpen is a transcription

regulator but may also be expressed on the surface of vascular endothelial cells. This hypothesis remains the subject of active investigation. Nonetheless, it is true that the aptamer could be used to specifically label microvessels surrounding rat glioblastoma, but not microvessels in normal, mature brain vasculature. As was the case with anti-PSMA aptamers, the anti-pigpen aptamers could be fluorescently labeled and used for contrast staining of transformed cells versus normal cells.

CONCLUSION

Aptamers have interesting properties relative both to protein therapeutics, such as antibodies, and to small organic drugs. Moreover, they possess qualities that make them great candidates for theoretical investigations. I will be using aptamers as the model of choice to better understand the implications behind the selection process. Many questions remain unanswered about this process. I will design selections in a way that will allow me to probe different starting points in sequence space and investigate whether the trajectory the molecules take correspond to each other. I will also investigate whether different starting points lead to different end points. The results of this study will shed light onto the question whether conventional *in vitro* selection process utilized is optimal for attaining the best binding species However, before this can take place, certain modifications to the automation system are essential. Issues of cross-contamination of selections using the same RNA pool against the same target will have to be addressed initially. This will not only aid in theoretical studies, but it is crucial for future selections against similar but more sensitive cellular targets using the same selection pools.

As the first aptamer therapeutics make their way into clinical practice uncertainties surrounding this class of therapeutic compounds will dissipate, and new aptamers should find increasing use as drugs, a progression that will be aided by new synthesis chemistries and new selection methodologies. Aptamers are slowly finding their way into cancer biology research as a diagnostics tool and site specific delivery [39-41]. Because of their increasingly important role in cancer research, and the importance of finding biomarkers for early detection, cell surface selection methods will be conducted.

Three lung tumor cell lines that differ in their Epidermal Growth Factor Receptor (EGFR) expression were chosen as targets for selection. NCI-H358 is from a non-small cell lung cancer and expresses wild type EGFR. NCI-1650 is from a bronchoalveolar carcinoma and expresses a mutant form of EGFR (in-frame deletion delE746-A750). NCIH526 is from a small cell lung cancer (SCLC) cell lines and expresses low to undetectable levels of EGFR. The first two lines are adherent, but NCIH526 is a suspension culture. Selection of aptamers against these three lines will serve several purposes: first, it will potentially demonstrate that aptamers against cell surfaces can be used to differentiate various types of lung tumor cells. Second, aptamers may be derived that recognize particular types or levels of EGFR, an important tumor marker. And finally, selection methods that target both adherent and suspension cells will be developed. Methods for tailoring specific aptamer for the targets of interest will be investigated and once an optimal method is obtained, it can be translated to the "contamination-free" automated platform that will be developed.

REFERENCES

1. **Ellington, A.D. and J.W. Szostak,** ***In vitro selection of RNA molecules that bind specific ligands.*** **Nature, 1990. 346(6287): p. 818-22.**

2. **Tuerk, C. and L. Gold,** ***Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.*** **Science, 1990. 249(4968): p. 505-10.**

3. **Blank, M. and M. Blind,** ***Aptamers as tools for target validation.*** **Curr Opin Chem Biol, 2005. 9(4): p. 336-42.**

4. **Proske, D., et al.,** ***Aptamers-basic research, drug development, and clinical applications.*** **Appl Microbiol Biotechnol, 2005. 69(4): p. 367-74.**

5. **Nimjee, S.M., C.P. Rusconi, and B.A. Sullenger,** ***Aptamers: an emerging class of therapeutics.*** **Annu Rev Med, 2005. 56: p. 555-83.**

6. **Yan, A.C., et al.,** ***Aptamers: prospects in therapeutics and biomedicine.*** **Front Biosci, 2005. 10: p. 1802-27.**

7. **DeAnda, A., Jr., et al.,** ***Pilot study of the efficacy of a thrombin inhibitor for use during cardiopulmonary bypass.*** **Ann Thorac Surg, 1994. 58(2): p. 344-50.**

8. **Bless, N.M., et al.,** ***Protective effects of an aptamer inhibitor of neutrophil elastase in lung inflammatory injury.*** **Curr Biol, 1997. 7(11): p. 877-80.**

9. **Bock, L.C., et al.,** ***Selection of single-stranded DNA molecules that bind and inhibit human thrombin.*** **Nature, 1992. 355(6360): p. 564-6.**

10. **Lehman, N., et al.,** ***The genotypic landscape during in vitro evolution of a catalytic RNA: implications for phenotypic buffering.*** **J Mol Evol, 2000. 50(5): p. 481-90.**

11. **Klug, S.J. and M. Famulok,** ***All you wanted to know about SELEX.*** **Mol Biol Rep, 1994. 20(2): p. 97-107.**

12. **Burgstaller, P., M. Kochoyan, and M. Famulok,** ***Structural probing and damage selection of citrulline- and arginine-specific RNA aptamers identify base positions required for binding.*** **Nucleic Acids Res, 1995. 23(23): p. 4769-76.**

13. **Mannironi, C., et al.,** ***Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif.*** **Rna, 2000. 6(4): p. 520-7.**

14. **Fontana, W. and P. Schuster,** ***Shaping space: the possible and the attainable in RNA genotype-phenotype mapping.*** **J Theor Biol, 1998. 194(4): p. 491-515.**

15. **Fontana, W., et al.,** ***RNA folding and combinatory landscapes.*** **Physical Review. E. Statistical Physics, Plasmas, Fluids, And Related Interdisciplinary Topics, 1993. 47(3): p. 2083-2099.**

16. **Schuster, P., et al.,** ***From sequences to shapes and back: a case study in RNA secondary structures.*** **Proc Biol Sci, 1994. 255(1344): p. 279-84.**

17. **Schultes, E.A. and D.P. Bartel,** ***One sequence, two ribozymes: implications for the emergence of new ribozyme folds.*** **Science, 2000. 289(5478): p. 448-52.**

18. **Held, D.M., et al.,** ***Evolutionary landscapes for the acquisition of new ligand recognition by RNA aptamers.*** **J Mol Evol, 2003. 57(3): p. 299-308.**

19. **Cox, J.C. and A.D. Ellington,** ***Automated selection of anti-protein aptamers.*** **Bioorg Med Chem, 2001. 9(10): p. 2525-31.**

20. **Sooter, L.J., et al., *Toward automated nucleic acid enzyme selection.* Biol Chem, 2001. 382(9): p. 1327-34.**

21. **Gschwind, A., O.M. Fischer, and A. Ullrich, *The discovery of receptor tyrosine kinases: targets for cancer therapy.* Nat Rev Cancer, 2004. 4(5): p. 361-70.**

22. **Hansford, J.R. and L.M. Mulligan, *Multiple endocrine neoplasia type 2 and RET: from neoplasia to neurogenesis.* J Med Genet, 2000. 37(11): p. 817-27.**

23. **Ichihara, M., Y. Murakumo, and M. Takahashi, *RET and neuroendocrine tumors.* Cancer Lett, 2004. 204(2): p. 197-211.**

24. **Cerchia, L., et al., *Neutralizing aptamers from whole-cell SELEX inhibit the RET receptor tyrosine kinase.* PLoS Biol, 2005. 3(4): p. e123.**

25. **Petitclerc, E., et al., *Integrin alpha(v)beta3 promotes M21 melanoma growth in human skin by regulating tumor cell survival.* Cancer Res, 1999. 59(11): p. 2724-30.**

26. **Mi, J., et al., *Targeted inhibition of alphavbeta3 integrin with an RNA aptamer impairs endothelial cell growth and survival.* Biochem Biophys Res Commun, 2005. 338(2): p. 956-63.**

27. **Guo, K., et al., *Aptamer-based capture molecules as a novel coating strategy to promote cell adhesion.* J Cell Mol Med, 2005. 9(3): p. 731-6.**

28. **Homann, M. and H.U. Goringer, *Combinatorial selection of high affinity RNA ligands to live African trypanosomes.* Nucleic Acids Res, 1999. 27(9): p. 2006-14.**

29. **Lorger, M., et al., *Targeting the variable surface of African trypanosomes with variant surface glycoprotein-specific, serum-stable RNA aptamers.* Eukaryot Cell, 2003. 2(1): p. 84-94.**

30. **Fitter, S. and R. James, *Deconvolution of a complex target using DNA aptamers.* J Biol Chem, 2005. 280(40): p. 34193-201.**

31. **Pan, W., et al., *Isolation of virus-neutralizing RNAs from a large pool of random sequences.* Proc Natl Acad Sci U S A, 1995. 92(25): p. 11509-13.**

32. **Cerchia, L., et al., *Nucleic acid aptamers in cancer medicine.* FEBS Lett, 2002. 528(1-3): p. 12-6.**

33. **Blank, M., et al., *Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels. selective targeting of endothelial regulatory protein pigpen.* J Biol Chem, 2001. 276(19): p. 16464-8.**

34. **Hicke, B.J., et al., *Tenascin-C aptamers are generated using tumor cells and purified protein.* J Biol Chem, 2001. 276(52): p. 48644-54.**

35. **Lupold, S.E., et al., *Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen.* Cancer Res, 2002. 62(14): p. 4029-33.**

36. **White, R., et al., *Generation of species cross-reactive aptamers using "toggle" SELEX.* Mol Ther, 2001. 4(6): p. 567-73.**

37. **Pinto, J.T., et al., *Prostate-specific membrane antigen: a novel folate hydrolase in human prostatic carcinoma cells.* Clin Cancer Res, 1996. 2(9): p. 1445-51.**

38. **Alliegro, M.C. and M.A. Alliegro, *Nuclear injection of anti-pigpen antibodies inhibits endothelial cell division.* J Biol Chem, 2002. 277(21): p. 19037-41.**

39. **Chu, T.C., et al., *Aptamer:toxin conjugates that specifically target prostate tumor cells.* Cancer Res, 2006. 66(12): p. 5989-92.**

40. **Chu, T.C., et al., *Labeling tumor cells with fluorescent nanocrystal-aptamer bioconjugates.* Biosens Bioelectron, 2006. 21(10): p. 1859-66.**

41. **Chu, T.C., et al., *Aptamer mediated siRNA delivery.* Nucleic Acids Res, 2006. 34(10): p. e73.**

CHAPTER 2: APTAMER DATABASE

INTRODUCTION

PUBLIC DATABASES AVAILABLE

The number of publicly available databases collectively constitutes a rich and ever-increasing source of biological information [1]. The information available is not limited to genomic sequences with different information obtained from experiments and computational experiments. The annotations include the location of introns, exons, and the existence of transcripts with splice variants. Databases that include protein-related information is also widespread and they contain valuable information about the 3D structures, base pair interactions, and other important information that is useful for any scholar. In order to facilitate data extraction by users, most databases are extensively cross-referenced and used highly efficient search tools that are developed mainly for sequence-based searches.

Some of the best known databases include the Gene Ontology (GO) Consortium, which contains information such as biological processes, molecular functions and cellular components [2]. Another useful database that focus on cellular processes and provide their information as interconnected pathways with computer representations is the Kyoto Encyclopedia Genes and Genome (KEGG) [3]. Other popular databases focus on microarray data [4, 5]. More specialized databases include those that focus on individual organisms such as such *D. melanogaster, A. Taliana* and *C. elegans*. [6-8].

THE APTAMER DATABASE

Functional nucleic acids can be selected from random sequence libraries. In general, *in vitro* selection mimics natural selection, in that a pool of heritable diversity is generated (typically by chemical synthesis). The pool is sieved for binding or catalysis, and successful variants are preferentially amplified by some combination of reverse transcription, PCR, and *in vitro* transcription [9-13]. This process has also been described as the Systematic Evolution of Ligands by EXponential enrichment or SELEX [14]. Nucleic acid binding species generated by *in vitro* selection have been referred to as aptamers [15]. Aptamers can be RNA, modified RNA, single-stranded DNA, or double-stranded DNA and have been selected to bind targets ranging from small organic molecules to entire organisms. Novel nucleic acid catalysts can also be selected, in general by modifying selection schemes so that variants that make or break covalent (rather than non-covalent) bonds are selectively retained in the population [16-19]. Since its introduction over ten years ago, *in vitro* selection has been widely adopted as a tool for the development of research reagents. This method shows promise for the generation of diagnostic and therapeutic agents [10, 20-22].

The Aptamer Database is not only extremely useful both for identifying what aptamers and unnatural ribozymes already exist, but also for garnering information about *in vitro* selection experiments as a whole and for better understanding the distribution of functional nucleic acids in sequence space and the topographies of fitness landscapes. We have collaborated with theoretical biologists for several years on analyses of the Aptamer Database, and now wish to make this resource much more widely available. In

addition, comparative sequence analysis tools should facilitate mappings between natural and unnatural sequence spaces, ultimately providing insights into both. For example, selected transcription factor binding sites have proven to be similar to and predictive of natural transcription factor binding sites [23-26].

Like other types of sequence data, the amount of sequence data generated by *in vitro* selection experiments has been accumulating exponentially. Because the sheer number and diversity of selection experiments has risen over the years; it is essential to gather all the sequence data into a comprehensive, continuously updated database. Unfortunately, Genbank and other sequence databases do not have extensive collections of non-natural sequences, and journals do not typically require the entry of non-natural sequences into these databases. We have now privately maintained the Aptamer Database for five years, and expanded its content on a monthly basis.

Another database, the SELEX_DB, also contains some information from *in vitro* selection experiments [27, 28]. However, the SELEX_DB focuses on *in vitro* selection experiments that have helped define natural DNA and RNA recognition sites for proteins, rather than including the entire repertoire of *in vitro* selection experiments. While there is some overlap between the two sites, the Aptamer Database is in general more complete, and contains entries from over 300 published *in vitro* selection experiments; in contrast SELEX_DB has entries from only 116 publications. The focus of the SELEX_DB on known binding sites ultimately limits its utility for exploring connections between selection experiments and the natural world. For example, natural aptamers that can bind ligands and regulate gene expression (so-called 'riboswitches') have been discovered by Breaker and his co-workers [29-31], and the sequences of riboswitches and

aptamers that both bind cyanocobalamin can be readily compared using the Aptamer Database.

DATABASE CONTENT

We wanted to create a database that would be easily accessible containing the essential information from *in vitro* selections carried out up-to-date. The Aptamer Database contains sequences drawn from over 300 published *in vitro* selection experiments. Each entry is described by the following fields: Author (last name, first name, and middle initial for each author); Title; Medline (accession number and a direct link to the Pubmed record); Target name (the name of the ligand that was used for the selection of an aptamer); Target type (the classifications we have chosen are proteins, peptides, nucleic acids, organic molecules, inorganic molecules or other); Journal (year, volume, issue, pages); Pool category (DNA or RNA); Modified (Y or N) (indicates whether the nucleic acid pool used in a selection was natural or modified); Buffer conditions; Template description (describes the length of the random region); Template sequence (describes the primer binding sites); and Sequences (the list of all the sequences isolated from the selection).

References to and sequences from *in vitro* selection experiments can be searched by providing specific queries relevant to one of the fields such as author name, target name, type of target, type of pool, and so forth, as shown in **Figure 2.1A.** In this example the database is being searched for a given Target, thrombin [32]. Once the user makes a selection, the results will show all the relevant papers that match the criteria supplied as

shown in **Figure 2.1B.** The user can then hone in on the information in one or more particular publications, as shown in **Figure 2.1C.**

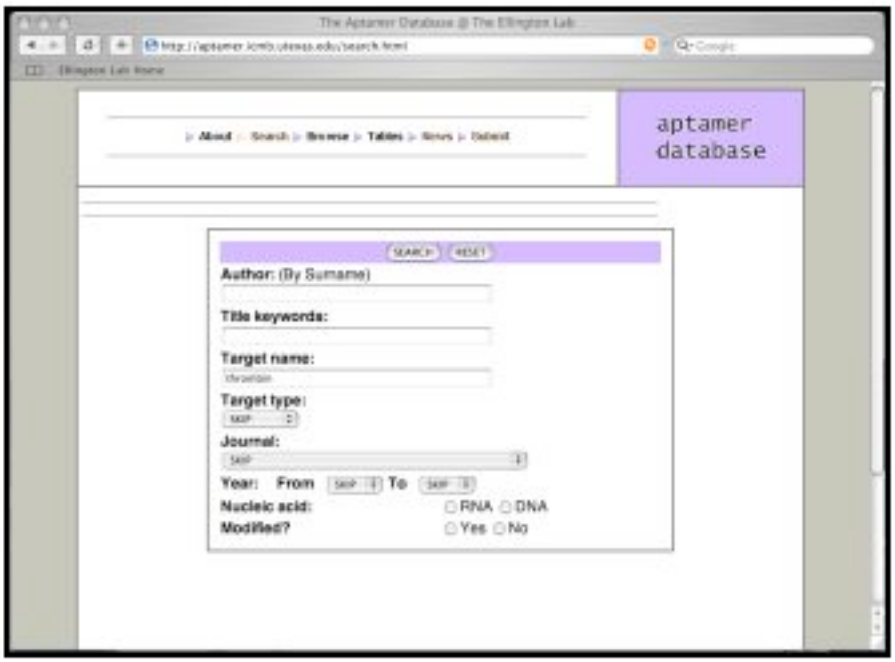


Figure 2.1A. Aptamer Search Page. Users can search the database by supplying a broad range of terms such as Author's last name, Title keywords, or Target name. Records in the database can also be searched based on combination of different criteria (such as the Target type, the journal or year a particular record was published, and the type of pool (RNA or DNA) that was used to carry out the selection). In the example shown, the term "thrombin" is supplied in the Target name dialogue box.

The database is updated monthly as new papers on the *in vitro* selection of aptamers or unnatural ribozymes become available. Initially, older papers were chosen for entry based on using the keywords "aptamer" and "SELEX" with the Pubmed search engine. Additional searches with the same keywords were also conducted using the SciFinder search engine (available at http://www.cas.org/SCIFINDER/scicover2.html).

Pubcrawler [33] is utilized for monthly updates using the keywords "aptamer" and "SELEX." At present, data is entered manually. Each sequence is entered in the database template twice and the entries are compared to ensure accuracy. While I have attempted to use optical character recognition (OCR) software, it has proven both inefficient and inaccurate, since the formats of published data are very different from one another.

Figure 2.1B. Results of the Search Page.. A display of the multiple records that were found for the target "thrombin."

While I have attempted to make the database as complete as possible, some selection papers have been published without appropriate or standard keywords, and thus that some literature may have been overlooked. I have included older references as they are

brought to my attention. In addition, one of the prime considerations in publishing the Aptamer Database is to facilitate the entry of data by authors and users, and have provided a template for data entry available for download at http://aptamer.icmb.utexas/submit) [34]. Only data that has been peer-reviewed in conjunction with a publication will be accepted in the database.

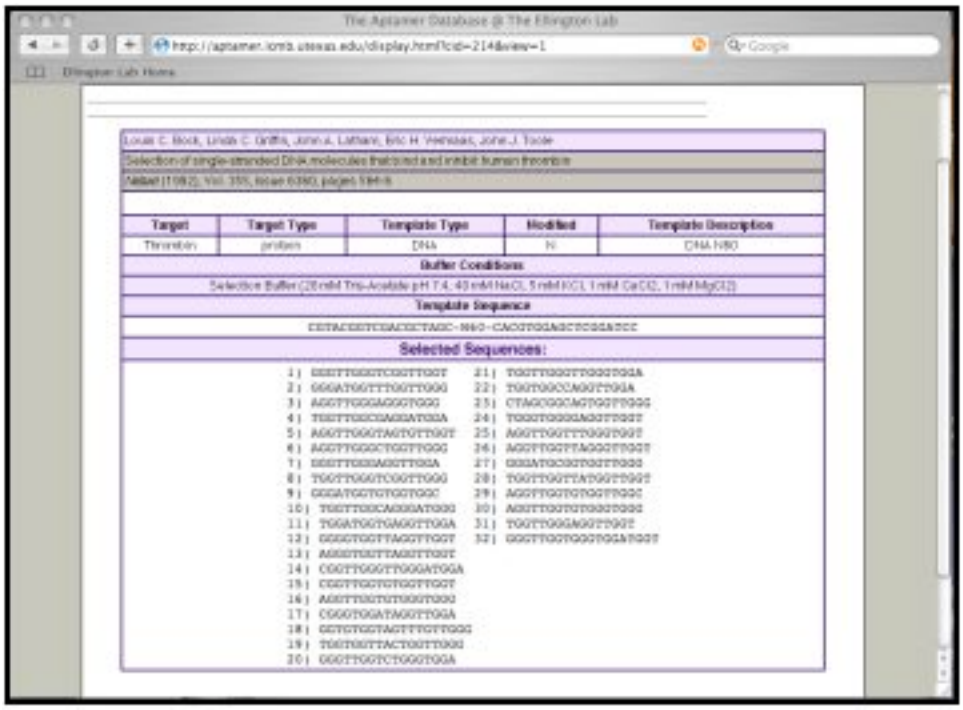


Figure 2.1C. Sequences from the Search. Any one of the recovered records has an associated set of aptamer or ribozyme sequences. For the "thrombin" example, the sequences from the original single-stranded DNA selection that targeted thrombin [32] are shown. There is also more detailed information about the selection that produced these sequences, such as the nature of the pool and buffer conditions used for selection.

DATABASE CONTENT DISTRIBUTION

Once the information was stored in the database, it was necessary to look into the distribution of the data that was available to the public. It allowed for a glimpse to the overall amount of *in vitro* selection data based on the type of selection that have been conducted. First step was to investigate the distribution of the different targets that have been selected. **Figure 2.2** shows the amount of different target-types that have aptamers selected. Almost half of the targets fall under the "protein" family. Other classifications include inorganic molecules, organic molecules, nucleic acids, peptides, and "other" which include whole organisms, whole cells, and others molecules that cannot be classified under the other existing categories.

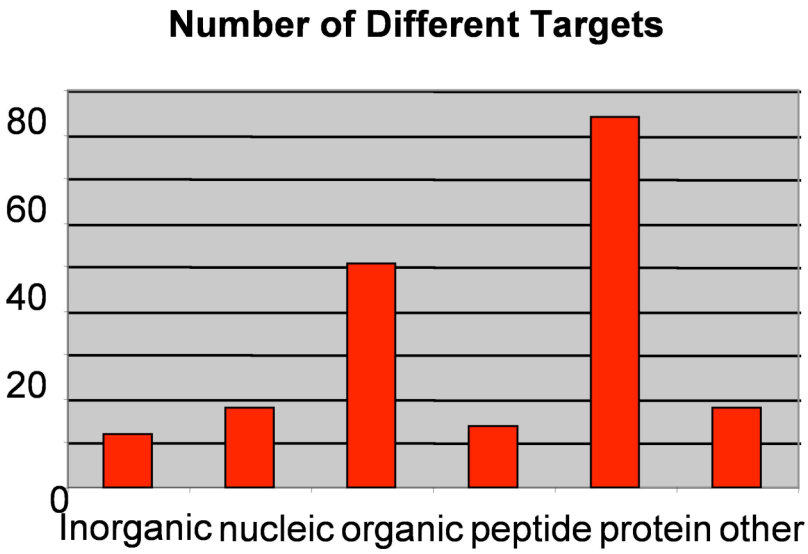


Figure 2.2. Distribution of Selected Targets that are Found in the Database. Most of the targets that have been utilized are proteins. Targets that fall under the "other" category include whole organisms and mammalian cells.

The next set of information of interest was the number of aptamers that were found in the database for each target category as shown in **Figure 2.3.** Once a clear understanding of the distribution of the aptamers was available, we wanted to investigate whether most of the aptamers that were selected were RNA or DNA. Interestingly, most of the selections that have been carried out are RNA aptamer selections as shown in **Figure 2.4**. They include modified RNA selections. The reason for the abundant number of RNA based selections can be attributed to the fact that the method for RNA selections was introduced before that of DNA. Also, for therapeutic purposes, RNA molecules do not have the risks associated with DNA such as genome integration into the host.

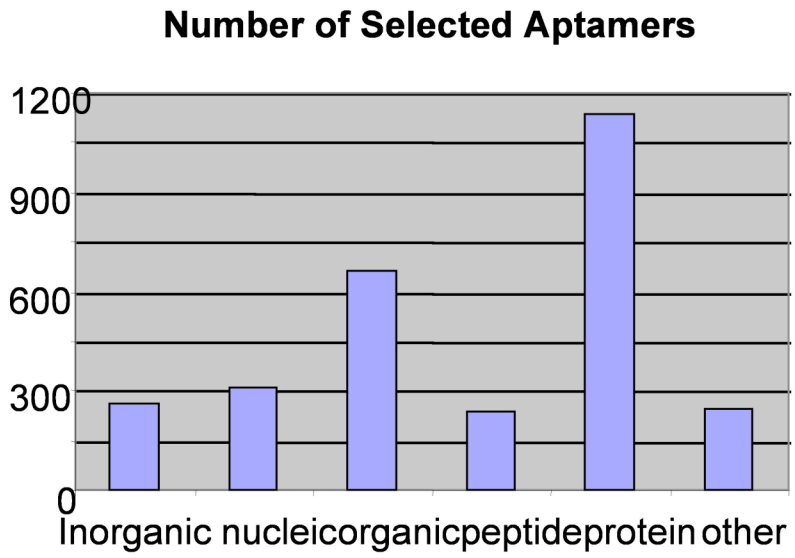


Figure 2.3. Distribution of Aptamers for each Target Group. As shown in the figure, there is more aptamers target to proteins that any other group. This can be attributed to the fact that the fraction of protein targets in the database is the highest.

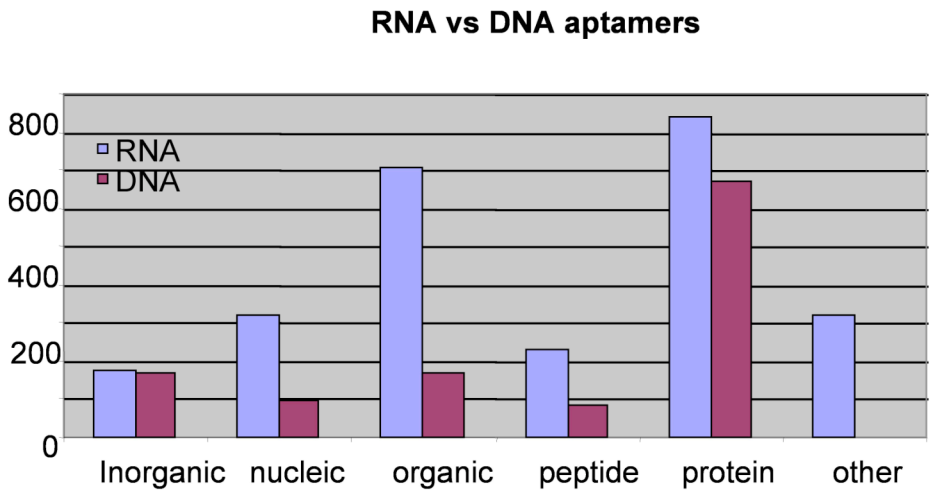


Figure 2.4. Distribution of RNA Versus DNA Aptamers. The number of RNA aptamers that have been selected is more abundant than of DNA aptamers. Many different factors may contribute to such a disparity such as the selection methods as well as the versatility of RNA aptamers.

CHAOS GAME REPRESENTATION OF THE APTAMER DATABASE

Once the distribution of RNA versus DNA aptamers was obtained. The next challenge was to represent the sequence data from the database in a clear and concise manner. One of the greatest challenges is visual data representation of vast amounts of sequence data. A very interesting way that we have found to represent data is known as the Chaos Game Representation (CGR). Jeffrey introduced CGR in 1990 as a tool for studying the "non-randomness" of genomic sequences [35]. It used primary sequence as data input and it would place the sequence within a quadrant based on its nucleotide content. **Figure 2.5** shows the CGR representation of all DNA and RNA aptamers irrespective of each other. The quadrant has four vertices each represented by one of the four nucleotides. Within the square images obtained by the method each aptamer is

represented by a dot. Sequences are placed based on their nucleotide composition. The CGR quadrants are graphed so that the lower (A:T) and upper (G:C) halves indicate the base composition and the diagonals indicate the purine/pyrimidine composition. Additional sub-quadrants can be further subdivided containing sequences ending with a given di-nucleotide. This will allow the sequences that differ only in one letter to be located in adjacent quadrants. The CGR representation depicts clearly that selected DNA aptamers were AT rich as opposed to RNA aptamers that seemed to have an even representation in terms of nucleotide composition. The advantage of conducting this type of graphical representation is that it does not depend on for sequence length, which is a challenge for comparing aptamer sequences.

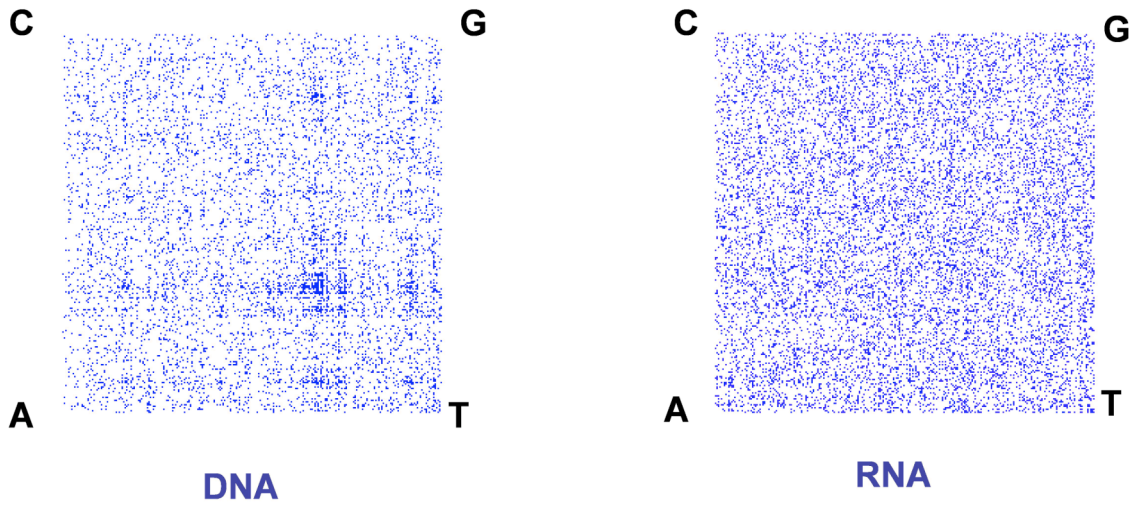


Figure 2.5. CGR Representation of Aptamers. Sequences from the database are extracted and divided into DNA and RNA aptamers. The nucleotide composition for each category is analyzed and projected. The analysis is based on an algorithm described by the Fletcher group in 2001 [36]. DNA aptamers are skewed towards AT rich as opposed to RNA aptamers that remains relatively uniform in ACTG composition.

GLOBAL REPRESENTATION OF THE APTAMER DATABASE

To get an overall distribution of all the sequences in the database, a space-filled model of aptamer position relative to each other was created. A global analysis of the aptamers was conducted to determine the distribution of the sequences in space. Using a Basic Local Alignment Search Tool (BLAST) algorithm, with the default settings, pairwise comparisons of all the aptamer sequences were conducted. The aptamers were positioned on the map according to sequence similarity. The results were visualized using a large graph layout (LGL) to visualize the extensive map summarizing the results of the sequence comparisons. The map served as theoretical framework that provided inferences about aptamer similarities. Aptamers that were selected against the same target were often adjacent to each other as shown in **Figure 2.6.**

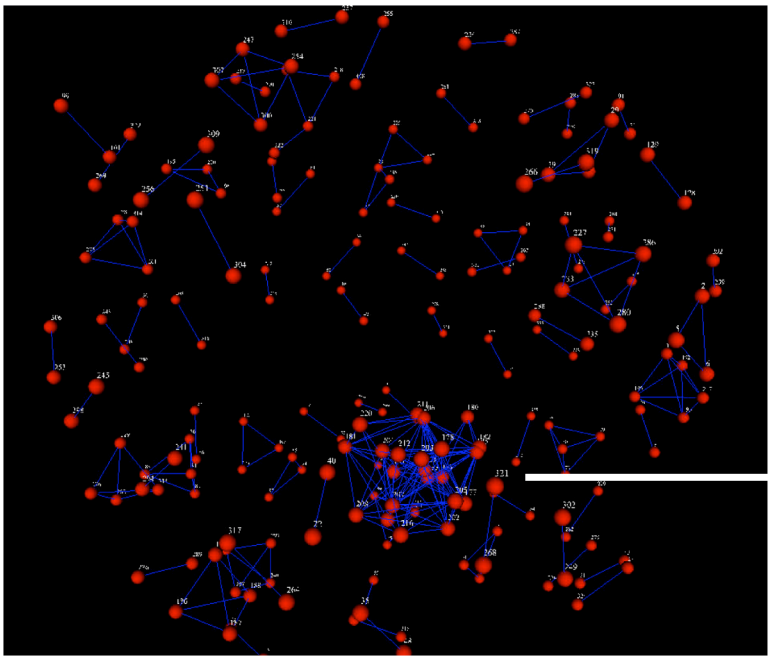


Figure 2.6. Global Distribution of RNA Aptamers in the Aptamer Database. A BLAST analysis of the RNA aptamer sequences in the database shows that aptamers that are isolated from the selections towards the same target were more closely related to each other than those selected against other targets. To construct this map, the aptamer sequences were compared each other using the BLAST software using default settings. A vertex represents one aptamer and a significant blast value is represented by an edge connecting the vertices. To reduce the multidimensionality of the BLAST relationships, an algorithm previously described by the Marcotte group is used [37].

SEQUENCE INFORMATION ANALYSIS

Given that aptamers and unnatural ribozymes were derived from random sequence libraries, we wanted to investigate whether functional sequences shared any qualities to random sequences and study the properties that constitute a functional sequence. This question is more than just academic, in that a bias in sequence composition or function could help to inform genomic searches for natural, functional RNA molecules [38-40].

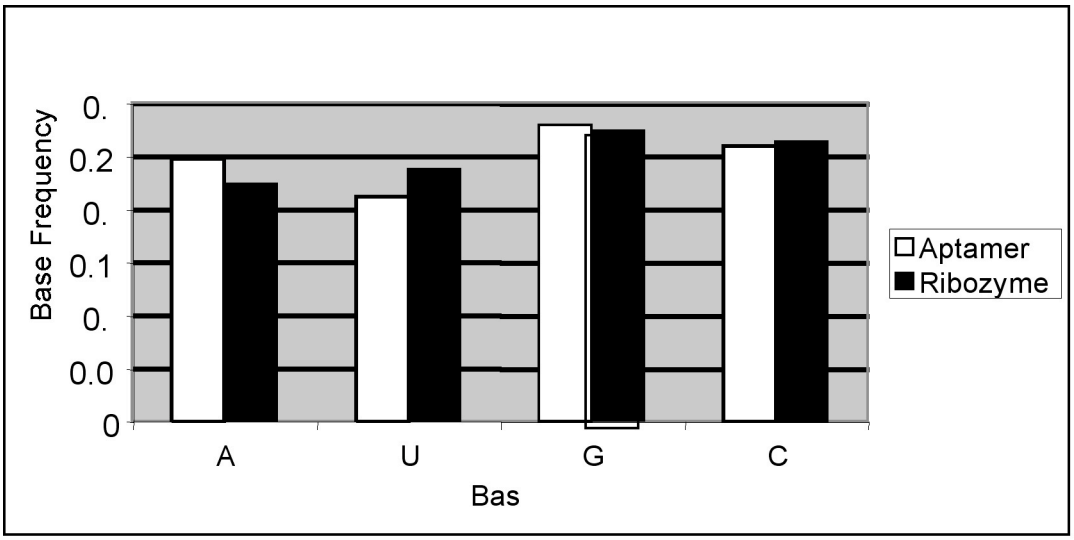


Figure 2.7 . Sequence analyses based on the Aptamer Database. The random regions of all RNA aptamer and ribozyme sequences from the Aptamer Database were extracted,

and aggregate base compositions were determined. White bars represent the nucleotide frequencies calculated for RNA aptamers. Black bars represent the nucleotide frequencies calculated for ribozymes. The individual residues (A, G, C, or U) are shown on the X-axis, while the Y-axis represents the frequency of each residue [34].

In this regard, in collaboration with the Professor Lauren Meyers, a base composition analysis of all of the sequences from the database reveals that there is a slight statistical skewing towards G and C (p-value < 0.01) in the base compositions of RNA aptamers and ribozymes relative to completely random sequences (equimolar A,C,G, and U; **Figure 2.7**).

As Schultes et al. [41] have previously observed, functional sequences, whether natural or unnatural, appear to have a slight preponderance of guanosine and cytidine. Interestingly, while the content of (G+C) appears to be similar for aptamers and ribozymes, ribozymes contain proportionately more U (and less A) than aptamers (p-value < 0.01). Beyond demonstrating that functional nucleic acid sequences of all sorts have particular sequence characteristics, these broad sequence analyses may inform the design of random sequence pools for selections; for example, it may be useful to skew the composition of a pool for the selection of ribozymes to a G:A:U:C ratio of 0.28:0.22:0.24:0.26. Since many selected ribozymes differ in sequence and function and cannot be aligned, this sort of analysis could be made much easier with the Aptamer Database.

The collected availability of sequences in the database also facilitates other global analyses. For example, we exported all of the RNA aptamer and ribozyme sequences and analyzed their potential for forming secondary structures using the program RNAfold from the Vienna RNA package [42]. The minimum free energy algorithm in RNAfold is

based on the dynamic programming algorithm developed by Zuker, et al. [43]. The Zuker algorithm only generates minimum free energy structures. Another alternative would have been to look at the ensemble of lowest energy structures. Lawrence and his co-workers conducted a study that compared the reliability of different RNA structure prediction algorithms [44]. Probability models that generate ensembles, such as McCaskill's algorithm [45], did prove to be a better predictor of the structures of longer sequences. However, for shorter (< 120 nucleotides) sequences, the improvements in structure prediction were modest and could primarily be attributed to the inability of the Zuker algorithm to take into account pseudoknots, a motif that is generally difficult to predict for any algorithm. In the end, the comparative analysis [46] confirmed that the Zuker minimum free energy method was very reliable for structure prediction. This method should be particularly appropriate for obtaining a general overview of the structural characteristics of the short selected sequences found in the database.

The fraction of paired nucleotides in folded aptamer and ribozyme structures were analyzed and assessed to see whether they were significantly different from the fraction of paired nucleotides in folded randomized structures. In this instance, the fraction of base pairs found in selected aptamer and ribozyme sequences is statistically greater than the fraction found in randomized versions of the same sequences as shown in **Figure 2.9A** (p-value < 0.01). In addition, in selected sequences, G:C pairings are more abundant than the A:U or G:U pairings, as shown in **Figure 2.9B** (p-value < 0.01). The large excess of G:C pairings is not a result of the only slightly higher concentrations of G and C in selected nucleic acids. Taken together, these results reveal that selection for binding or catalytic function of necessity results in selection for secondary structural stability.

Schultes et al. [41] have also shown that there is a preponderance of G:C pairings in the stem regions of natural, functional RNA molecules relative to unpaired regions such as loops. Similarly, examinations of natural sequences have suggested that G:C pairings are required to stabilize and present protein-binding sites [47-49].

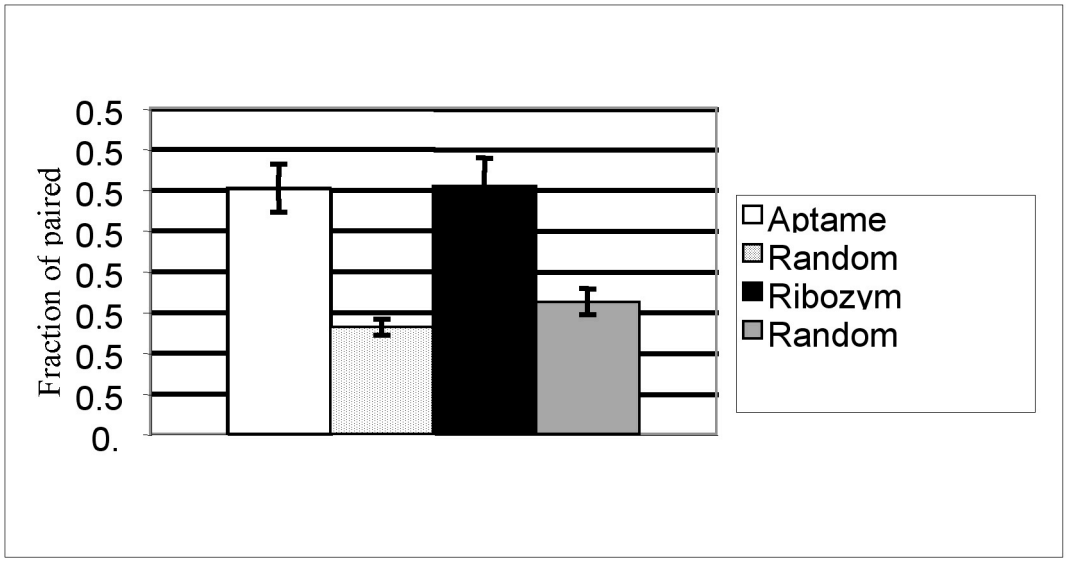


Figure 2.8A. Base-pairing in selected sequences. Overall base-pairing in RNA aptamers and ribozymes sequences. The fractions of base-paired residues were determined for RNA aptamers (white bar) or ribozymes (black bar). To determine whether the number of base-pairs that were formed was significantly influenced by selection, the random regions were randomized and re-folded for the RNA aptamers and ribozymes [34].

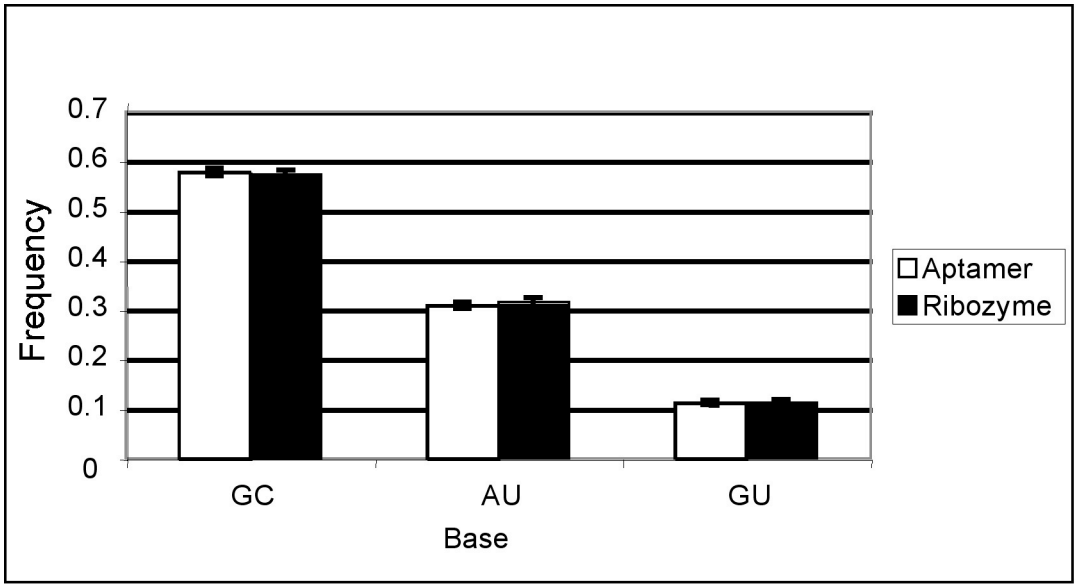


Figure 2.8B. Distribution of Individual Base Parings. The data from **Figure 2A** was re-tabulated in terms of the types of base-pairs formed [34]. The white bar represents RNA aptamer sequences, and the black bar represents ribozymes sequences. The Y-axis represents the fraction of base-pairs in each class, normalized to the total number of base-pairings found in **Figure 2A**.

ROBUSTNESS OF SECONDARY STRUCTURES

More complex analysis has been conducted based on the data acquired from the Aptamer Database. In collaboration with Dr. Lauren Meyers and Matt Copperthwaite, an analysis comparing the robustness of natural and unnatural secondary structures was carried out [50]. The premise of this analysis is that sequences of functional molecules have more stable secondary structures than random sequences. The question was whether this theory holds with equal importance for both artificially and naturally occurring molecules. Over 1000 evolved molecules were compared using two levels of structural

resolution: (i) The coarse-grained model, which uses the positions of structural components such as hairpins, stacks, or multiloops; and (ii) the standard secondary structure model that takes into account every nucleotide pairing position.

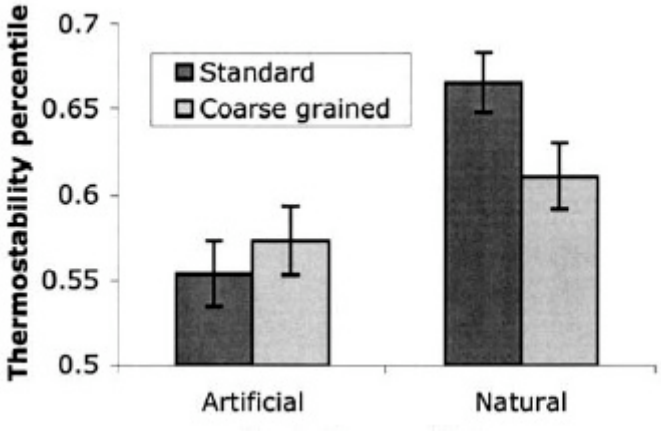


Figure 2.9. Comparative Thermostability of Artificially and Naturally Evolved Molecules. Average percentile thermostability for artificial molecules (aptmers) and natural molecules (nonhuman eukaryotic tRNA; human tRNA; archaebacterial 5S rRNA; eubacterial 5S rRNA; eukaryotic 5S rRNA; natural hammerhead ribozymes). We use the Boltzman coefficient of the minimum free energy shape to estimate thermostability. The percentile is the rank of a sequence's thermostability compared to the thermostabilities of 1000 randomly generated sequences of the same length and base composition. Percentiles significantly above 0.5 are more thermostable than expected from a random sequence.

Figure 2.9 [50] shows that the thermodynamic stability of those naturally occurring molecules are higher than those of artificially selected molecules (aptamers). The notion that directed evolution is comparable to natural evolution is challenged based on these results. The level of optimization of the structures is not comparable to that perfected from years of natural selection. A similar trend is shown when analysis of primary sequences is carried out in chapters 3 and 4. Further explanations about the potential reason as to for the more stable secondary structures not being selected are explained in later chapters.

CONCLUSIONS

The sequence and structural comparisons that have been carried out were relatively simplistic, they justify the contention that nucleic acids selected *in vitro* possess attributes similar to those of sequences found in nature. The Aptamer Database should continue to be a helpful source for such comparisons. For example, as new sequence or structural motifs are found, such as the well-characterized tetraloop sequences that have been found to stabilize natural RNA sequences [46, 51, 52], their prevalence and utility can be further confirmed by an unbiased search of the Aptamer Database.

METHODS AND MATERIALS

APTAMER DATABASE

The current version of the Aptamer Database contains entries from 237 papers that describe the *in vitro* selection of aptamers, ribozymes, and deoxyribozymes. The database primarily contains catalysts that have been selected from completely random pools, as opposed to variants on known, natural ribozymes. Each entry is described by the following fields: Author (last name and first name of the authors of each publication); Title; Medline Accession Number (allowing a direct link to the Pubmed record); Target (name of the ligand that was used for selection); Target Type (currently classified into proteins, peptides, nucleic acids, organic molecules, inorganic molecules or other);

Journal (year, volume, issue, pages); DNA/RNA or Modified (indicates whether the initial nucleic acid pool used for the selection was 'natural' or contained modified nucleotides); Buffer Conditions; Template Description (describes the length of the random region); Template Sequence (describes the sequences of the terminal primer binding sites); Sequences (list of each aptamer or ribozyme sequence that was selected, not including the Template Sequence or constant region). The database is updated monthly. Data is entered manually into the database. Each sequence is entered twice and the two entries are compared to ensure accuracy. The Aptamer Database is publicly available through http://aptamer.icmb.utexas.edu. To facilitate data entry by other users, a template of the database is available for download at http://aptamer.icmb.utexas/submit.

A database management system (DBMS) is necessary to set up and maintain the database. Different DBMSs exist, but for relational databases MySQL and POSTGRESQL are the preferred DBMSs for their speed and ease of use. Both are publicly available without licensing fees for educational institutions and have extensive technical support from the community. All the information in available in the database is store in a MySQL database server. Information is provided as a text file and a Perl pipeline has a modular structure that fills the tables of the MySQL database.

A Web Server is crucial for the database to be publicly accessible. A server that allows the generation of dynamic web pages and the ability to interpret and execute scripts is necessary. Apache is a web server that is freely available with the required capabilities. Typically scripts written in Perl or PHP are used as the backend database interface language and they follow the same licensing as previously described. Software that will translate and display specific contents of the database to the Internet for public

use will be needed. Many popular web interface software are available. The choice will be dependent on the complexity of the web page and the functionalities the developer is trying to attain. The database's web interface is based on the PHP language, and manages all the incoming queries. All available publication was found using a keyword search in literature databases such as Pubmed, Scifinder, and Chemical Abstracts.

GRAPHICAL REPRESENTATION

To construct the Chaos Game Representation (CGR) of the database, all the sequences were extracted in text format and the nucleotide composition of each sequence was analyzed. Only the random regions of the aptamers were accounted for during this analysis. The analysis is based on the algorithm presented by the Fletcher group in 2001 [36]. In summary, the CGR space was generated by the aptamer sequences is a planar illustration confined by the four possible nucleotides as vertices of a binary square.

To construct the map and space filled model of the database, we compared the nucleotide sequences with each other using the program BLASTN using default settings. The results of these were represented as a vertex in a network, and each significant BLAST similarity (E-scores of less than 10^{-9}) was represented as an edge connecting the corresponding aptamers. The network was then converted to an undirected network by creating a single edge between two connected sequences and retaining the more significant of the two BLAST E-values as the weight. Once the coordinates of the layout were obtained, they are fed into LGL, which is an algorithm introduced by the Marcotte group used to visualize very large biological networks [37].

LGL is useful for visualizing the results nucleic acid sequence comparisons, made using the program BLAST. The results are interpreted as a large biological network, where each aptamer is represented as a vertex, and each significant BLAST similarity is represented as an edge connecting the vertices. This general approach is known to effectively identify both close and distant sequence homologs.

SEQUENCE INFORMATION ANALYSIS

The base composition analysis is carried out using a Perl script that would count the number of each nucleotide frequency for aptamer and ribozymes. Secondary structures of the random regions of RNA aptamer and ribozyme sequences were obtained using the program RNAfold, implemented in the Vienna RNA Package [42]. The minimum free energy algorithm was based on the dynamic programming algorithm developed by Zuker et al. [43]. The fractions of base-paired residues were determined for RNA aptamers (white bar) or ribozymes (black bar). The base-pair fractions for all individual aptamers or all individual ribozymes were averaged; the error bars provide an indication of the spread of these values. To determine whether the number of base-pairs that were formed was significantly influenced by selection, the random regions were randomized and re-folded for the RNA aptamers and ribozymes.

MEASURING THERMODYNAMIC STABILITY

We computationally estimated the extent to which single molecules are buffered against thermodynamic noise. First, we computed the repertoire of structures that are

accessible to a sequence and approximated the probability of each structure. We used an extension of the standard thermodynamic minimum free energy folding algorithm, which permits the computation of all secondary structures within some energy range above the minimum free energy [42]. This algorithm provides reasonable approximations but is by no means perfect. It does not predict pseudoknots or other tertiary interactions, which are known to occur in both natural and artificial RNA molecules. The parameters are estimated at the physiological temperature of 37 C, which is appropriate for prediction of natural sequences. Artificial RNAs, however, are often selected and optimized at a lower temperature (25 C). These algorithms may thus make less accurate predictions for artificially selected molecules. Despite these limitations, we used the suboptimal folding algorithm to rapidly approximate the low energy portion of the secondary structure space of a given sequence. We neglected energy barriers and assumed that a sequence equilibrates among all structures whose free energy is within 5 kT of the groundstate. The 5-kT choice amounts to approximately 3 kcal at 37 C and corresponds to the loss of two CG/GC stacking interactions. Under thermodynamic equilibration, we assumed that the Boltzmann probability of a shape S, $\exp(-\Delta G_s/kT)/Z$, approximates the overall fraction of time that the molecule spends in s, where ΔG_s is the free energy of structure S, k is the Boltzmann constant, T the absolute temperature, and $Z = \sum \exp(-\Delta G_s = kT)$ is the partition function, which is computed by an algorithm previously described [45]. We used the Boltzmann probability of the lowest free energy state (the groundstate) to estimate the thermodynamic stability of a molecule. Thus we did not use minimum free energy alone to estimate thermostability but, instead, considered the extent to which a groundstate is stabilized with respect to competing shapes. In measuring thermodynamic

stability, we assayed secondary structures at two levels of resolution. The coarse grained secondary structure of a molecule notes the relative position of the following structural components: hairpins (H), interior loops (I), bulges (B), multiloops (M), and stacks (S). We first predicted the standard structure of a molecule, and then parsed the standard representation into these five components. When several suboptimal standard shapes reduced the same coarse-grained shape, we grouped them together and summed their Boltzmann coefficients. The coarse grained groundstate was the shape with the largest (collective) Boltzmann, which may or may not have corresponded with the original standard groundstate. In essence, the standard secondary structure looks at the precise sequence and structure of a molecule, while the coarse-grained structure is a measure of the overall fold, independent of sequence. By examining coarse-grained structures, we grouped structurally similar molecules. This may be appropriate, for example, if two nucleic acids are functionally similar because of a shared stacked helical junction, but these junctions in the two molecules differ in the number of single-stranded residues separating the two participating helices. Such molecules would be substantially different in the standard representation but appropriately equivalent in the coarse-grained representation. Coarse-grained structures also allow for some degree of error in the thermodynamic predictions and, thereby, provide an additional check on the validity of the analysis.

REFERENCES

1. **Galperin, M.Y., *The Molecular Biology Database Collection: 2005 update.* Nucleic Acids Res, 2005. 33(Database issue): p. D5-24.**

2. **Harris, M.A., et al., *The Gene Ontology (GO) database and informatics resource.* Nucleic Acids Res, 2004. 32(Database issue): p. D258-61.**

3. **Kanehisa, M., et al., *The KEGG resource for deciphering the genome.* Nucleic Acids Res, 2004. 32(Database issue): p. D277-80.**

4. **Brazma, A., et al., *Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.* Nat Genet, 2001. 29(4): p. 365-71.**

5. **Parkinson, H., et al., *ArrayExpress--a public repository for microarray gene expression data at the EBI.* Nucleic Acids Res, 2005. 33(Database issue): p. D553-5.**

6. **Chen, N., et al., *WormBase: a comprehensive data resource for Caenorhabditis biology and genomics.* Nucleic Acids Res, 2005. 33(Database issue): p. D383-9.**

7. **Gelbart, W.M., et al., *FlyBase: a Drosophila database. The FlyBase consortium.* Nucleic Acids Res, 1997. 25(1): p. 63-6.**

8. **Rhee, S.Y., et al., *The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community.* Nucleic Acids Res, 2003. 31(1): p. 224-8.**

9. **Burgstaller, P., A. Jenne, and M. Blind, *Aptamers and aptazymes: accelerating small molecule drug discovery.* Curr Opin Drug Discov Devel, 2002. 5(5): p. 690-700.**

10. **Brody, E.N. and L. Gold, *Aptamers as therapeutic and diagnostic agents.* J Biotechnol, 2000. 74(1): p. 5-13**

11. **Eaton, B.E., *The joys of in vitro selection: chemically dressing oligonucleotides to satiate protein targets.* Curr Opin Chem Biol, 1997. 1(1): p. 10-6.**

12. **Famulok, M. and G. Mayer, *Aptamers as tools in molecular biology and immunology.* Curr Top Microbiol Immunol, 1999. 243: p. 123-36.**

13. **Feigon, J., T. Dieckmann, and F.W. Smith, *Aptamer structures from A to zeta.* Chem Biol, 1996. 3(8): p. 611-7.**

14. **Klug, S.J. and M. Famulok, *All you wanted to know about SELEX.* Mol Biol Rep, 1994. 20(2): p. 97-107.**

15. **Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands.* Nature, 1990. 346(6287): p. 818-22.**

16. **Illangasekare, M. and M. Yarus, *A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis.* Rna, 1999. 5(11): p. 1482-9.**

17. **Huang, F., Z. Yang, and M. Yarus, *Self-capping RNA catalysts derived from selection-amplification.* Biol Bull, 1999. 196(3): p. 320-1.**

18. **Unrau, P.J. and D.P. Bartel, *RNA-catalysed nucleotide synthesis.* Nature, 1998. 395(6699): p. 260-3.**

19. **Ekland, E.H., J.W. Szostak, and D.P. Bartel, *Structurally complex and highly active RNA ligases derived from random RNA sequences.* Science, 1995. 269(5222): p. 364-70.**

20. **Lupold, S.E., et al., *Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen.* Cancer Res, 2002. 62(14): p. 4029-33.**

21. **Martell, R.E., J.R. Nevins, and B.A. Sullenger, *Optimizing aptamer activity for gene therapy applications using expression cassette SELEX.* Mol Ther, 2002. 6(1): p. 30-4.**

22. **Jayasena, S.D., *Aptamers: an emerging class of molecules that rival antibodies in diagnostics.* Clin Chem, 1999. 45(9): p. 1628-50.**

23. **Roulet, E., et al., *High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites.* Nat Biotechnol, 2002. 20(8): p. 831-5.**

24. **Kunsch, C., S.M. Ruben, and C.A. Rosen, *Selection of optimal kappa B/Rel DNA-binding motifs: interaction of both subunits of NF-kappa B with DNA is required for transcriptional activation.* Mol Cell Biol, 1992. 12(10): p. 4412-21.**

25. **Lebruska, L.L. and L.J. Maher, 3rd, *Selection and characterization of an RNA decoy for transcription factor NF-kappa B.* Biochemistry, 1999. 38(10): p. 3168-74.**

26. **Wright, W.E. and W.D. Funk, *CASTing for multicomponent DNA-binding complexes.* Trends Biochem Sci, 1993. 18(3): p. 77-80.**

27. **Ponomarenko, J.V., et al., *SELEX_DB: a database on in vitro selected oligomers adapted for recognizing natural sites and for analyzing both SNPs and site-directed mutagenesis data.* Nucleic Acids Res, 2002. 30(1): p. 195-9.**

28. **Ponomarenko, J.V., et al., *SELEX_DB: an activated database on selected randomized DNA/RNA sequences addressed to genomic sequence annotation.* Nucleic Acids Res, 2000. 28(1): p. 205-8.**

29. **Winkler, W.C., et al., *An mRNA structure that controls gene expression by binding S-adenosylmethionine.* Nat Struct Biol, 2003.**

30. **Winkler, W.C., S. Cohen-Chalamish, and R.R. Breaker, *An mRNA structure that controls gene expression by binding FMN.* Proc Natl Acad Sci U S A, 2002. 99(25): p. 15908-13.**

31. **Winkler, W., A. Nahvi, and R.R. Breaker, *Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression.* Nature, 2002. 419(6910): p. 952-6.**

32. **Bock, L.C., et al., *Selection of single-stranded DNA molecules that bind and inhibit human thrombin.* Nature, 1992. 355(6360): p. 564-6.**

33. **Hokamp, K. and K. Wolfe, *What's new in the library? What's new in GenBank? let PubCrawler tell you.* Trends Genet, 1999. 15(11): p. 471-2.**

34. **Lee, J.F., et al., *Aptamer database.* Nucleic Acids Res, 2004. 32(Database issue): p. D95-100.**

35. **Jeffrey, H.J., *Chaos game representation of gene structure.* Nucleic Acids Res, 1990. 18(8): p. 2163-70.**

36. **Almeida, J.S., et al., *Analysis of genomic sequences by Chaos Game Representation.* Bioinformatics, 2001. 17(5): p. 429-37.**

37. **Adai, A.T., et al., *LGL: creating a map of protein function with an algorithm for visualizing very large biological networks.* J Mol Biol, 2004. 340(1): p. 179-90.**

38. **Sudarsanam, P., Y. Pilpel, and G.M. Church, *Genome-wide co-occurrence of promoter elements reveals a cis-regulatory cassette of rRNA transcription motifs in Saccharomyces cerevisiae.* Genome Res, 2002. 12(11): p. 1723-31.**

39. **Hughes, J.D., et al., *Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae.* J Mol Biol, 2000. 296(5): p. 1205-14.**

40. **Klein, R.J., Z. Misulovin, and S.R. Eddy, *Noncoding RNA genes identified in AT-rich hyperthermophiles.* Proc Natl Acad Sci U S A, 2002. 99(11): p. 7542-7.**

41. **Schultes, E.A. and D.P. Bartel, *One sequence, two ribozymes: implications for the emergence of new ribozyme folds.* Science, 2000. 289(5478): p. 448-52.**

42. **Hofacker, I., et al., *Fast Folding and Comparison of RNA Secondary Structures.* Monatsh. Chem., 1994. 125: p. 167-188.**

43. **Zuker, M. and P. Stiegler, *Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information.* Nucleic Acids Res, 1981. 9(1): p. 133-48.**

44. **Ding, Y. and C.E. Lawrence, *Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond.* Nucleic Acids Res, 2001. 29(5): p. 1034-46.**

45. **McCaskill, J.S., *The equilibrium partition function and base pair binding probabilities for RNA secondary structure.* Biopolymers, 1990. 29(6-7): p. 1105-19.**

46. **Williams, D.J. and K.B. Hall, *Experimental and computational studies of the G[UUCG]C RNA tetraloop.* J Mol Biol, 2000. 297(5): p. 1045-61.**

47. **Varani, G., *Exceptionally stable nucleic acid hairpins.* Annu Rev Biophys Biomol Struct, 1995. 24: p. 379-404.**

48. **Morgan, W.D., D.G. Bear, and P.H. von Hippel, *Rho-dependent termination of transcription. I. Identification and characterization of termination sites for transcription from the bacteriophage lambda PR promoter.* J Biol Chem, 1983. 258(15): p. 9553-64.**

49. **Morgan, W.D., et al., *RNA sequence and secondary structure requirements for rho-dependent transcription termination.* Nucleic Acids Res, 1985. 13(10): p. 3739-54.**

50. **Meyers, L.A., et al., *The robustness of naturally and artificially selected nucleic acid secondary structures.* J Mol Evol, 2004. 58(6): p. 681-91.**

51. **Sorin, E.J., et al., *RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop.* J Mol Biol, 2002. 317(4): p. 493-506.**

52. **Li, W., B. Ma, and B. Shapiro, *Molecular dynamics simulations of the denaturation and refolding of an RNA tetraloop.* J Biomol Struct Dyn, 2001. 19(3): p. 381-96.**

CHAPTER 3: HIGH THROUGHPUT *IN VITRO* SELECTIONS OF APTAMERS

INTRODUCTION

ROLE OF AUTOMATION IN APTAMER SELECTIONS

Experimental procedures using automation systems have become widely accepted in modern science. With the increasing need to generate vast amounts of data for the fields of genomics and combinatorial chemistry, it almost intractable to produce such amounts of data with the conventional bench science. The improvements in this area have given way to robotically conducted experimental procedures to take over the repetitive practices of different experimental methods such as amplification reactions (PCR), sequencing reactions, microarray printing to name a few [1-3]. The integration of automation systems has eased the implementation of large-scale projects. Few examples of successful projects completed using robotic systems include the sequencing of the human genome, the mass spectroscopy analysis for protein-ligand interactions, and mammalian cell-line preparation [4-7].

It is not abnormal nowadays to see conventional biological labs equipped with robotic platforms to undertake tedious and repetitive tasks, which in turn frees up the end user with more time to conduct the complex experimental tasks and data analysis. The degree of automated implementation has become so advanced that the King group has designed robotic systems that are able to devise and test hypothesis and continually refined the experiments to meet the hypothesis [3]. An important advantage of automated

platforms is the consistency of the experimental procedures. This fact ensures that variations in the outcomes of certain procedures are indeed due to the nature of the reaction and not because of the manual manipulations by different scientists. Precision and consistency is highly sought after for scientific experiments. It is difficult to assess the results of *in vitro* selections when they are carried out manually since one cannot be sure whether the outcomes are due to discrepancies during the execution of the experiment. Since *in vitro* selections involve such a large amount of varied steps, robotic selections ensures the analysis of the results with higher confidence due to the reproducibility benchmarks that automation platforms provide. Previous students in the Ellington lab, Colin Cox and Letha Sooter have both perfected the *in vitro* selection processes to robotic platforms with a high degree of success [8-11].

EXISTING AUTOMATED SELECTIONS

While most of the selection protocols on robotics have been based on RNA pools [8-10], selections using ssDNA and dsDNA pools have been demonstrated as well [11]. Moreover, automated selections have been carried out within the lab using modified nucleotide pools. Translating *in vitro* selections to robotic systems, many steps in bench selections are omitted during automated selections. These steps include PAGE nucleic acid purifications; instead the focus went on to ensure efficient amplification chemistry by maximizing the generation of binding species and minimizing the generation of other products such as primer dimers or parasites. Selections to date have been single target selections with one pool, or selections of multiple different targets conducted simultaneously on the platform, or selections against multiple targets with different

nucleic acids pool with different primer sets. No simultaneous selections against the same target with the same nucleic acid pool have been conducted. The reason for this fact can be attributed to the challenge at eliminating contamination. Robotic platforms are open to their local environment. Encasements are available for some automations platforms; however, they are costly and a challenge in of itself to encase the platform that has third party hardware incorporated.

CHALLENGES OF AUTOMATED SELECTIONS

When I first embarked on performing automated selections, along with a former undergraduate student Travis Bayer, we observed that a certain degree of contamination occurred when selections were conducted in parallel when nucleic acids pools with the same primer binding regions were used. It appeared that the proximity of targets within a 96-well plate was enough for amplicons to contaminate neighboring wells. It was obvious that further modifications would be needed to overcome this problem. This would require a highly systematic approach to narrow the potential areas of contamination that can occur during the process of selection on an automated platform. For those who are interested in selecting against different targets or use different nucleic acid pools, this might seem like a minor problem. But for complex questions concerning evolutionary theories, this can certainly undermine the results of data collected. I have undertaken the task to assess whether contamination issues arise when parallel selections that are conducted with different pools that contain identical primer binding regions and selected against the same protein target. The significance of this investigation is to further give credibility to automation, especially when selections against very similar targets are to be

conducted simultaneously.

RESULTS AND DISCUSSION

NEW HOUSEKEEPING PROTOCOL

One of the main protocols that has been introduced into automated selections is the manipulating the initial pool in a "Round 0" hood free of contamination. The manipulation of every newly synthesized pool is confined inside the hood. Before and after each use, the hood is wiped down with bleach and "DNAse" Zap solution and "RNAse" Zap solution (Applied biosystems, Foster City, CA). In addition, after the hood is cleaned, UV radiation is used to further clean the surfaces. Except for UV radiation, robotic surfaces are meticulously wiped down with the same solutions.

Another great contributor to the contamination is the thermalcycler due to the lack of cover for the 96 well plate. We tested different sealing pads for the thermalcycler lid until we found an MJ "P" that allows an airtight sealing of the lid to the PCR plate so that during amplification no amplicons are able to escape the wells. The amplification steps have also been modified so that the last step of the process involving the temperature to equilibrate to room temperature for 10 minutes to reduce evaporation when the thermalcycler lid is opened.

INDIVIDUAL POOL AMPLIFICATION

Before any action was taken to amplify an intact pool, which contains approximately $1x10^{15}$ unique nucleic acid sequence, it was split into 8 different aliquots each containing approximately $1x10^{14}$ unique molecules to each aliquot. Calculating the probability of finding a sequence in a pool of sequences requires two pieces of information:

the probability of finding the sequence in a single attempt, and the number of attempts to find it. For a sequence with a random region of length n=40 where each position has 25% of containing A, G, T, or C; the probability of finding a certain sequence is:

$$P(\text{sequence}) = (1/4)^{40} = 8.3\text{E-}25$$

The probability of finding this sequence twice is P^2.= 6.8E-49. Each aliquot is individually large-scale amplified and transcribed into RNA. Eighteen rounds of selections were performed for the 8 pools against the protein target hen egg-white lysozyme simultaneously and analyzed as shown in **Figure 3.1.**

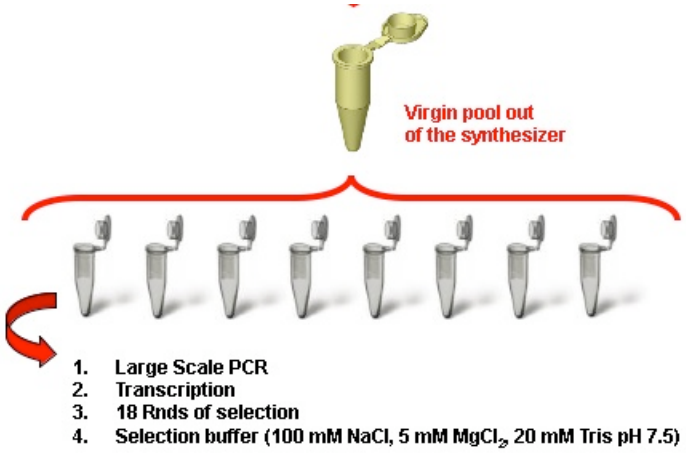


Figure 3.1. Initial Preparation of Newly Synthesized Pool. The newly synthesized "virgin" pool is subdivided into 8 different aliquots before amplification steps have taken place. This ensures that each aliquot will be exclusive distinct from each other.

I started my selections with a newly synthesized N30 random sequence pool that was previously designed by former students Colin Cox and optimized by Sulay Jhaveri from the laboratory. The N30 pool contains a random region flanked by two constant regions. **Figure 3.2** depicts the N30 pool design.

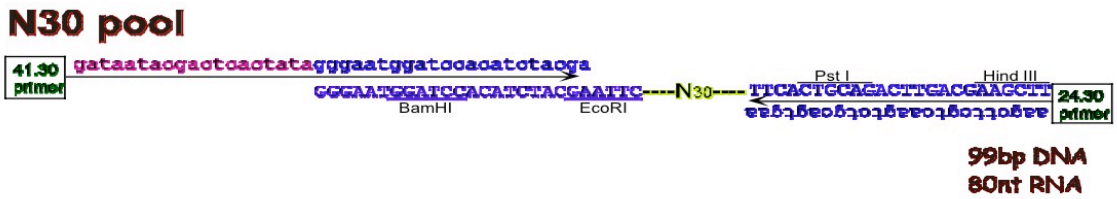


Figure 3.2. N30 Pool Representation. The bases that subsequently transcribed into RNA are shown as capitalized letters. Primer binding sequences are shown in lower case letters. The 5' prier 41.30 incorporates a T7 promoter "gataatacgactcactata." This pool has been previously designed to contain restriction endonuclease recognition sites, which are underlined.

The N30 pool's dynamics are very well known, and multiple selections have been conducted in the Ellington Laboratory using this pool [9, 10]. However, due to the widespread use of such pool for *in vitro* selections both at the bench and on robotic platforms, there was a high degree of a dominant contaminant, an aptamer clone that was previously selected against the same target Lysozyme [8].

To overcome the contamination issue, a second pool that was subsequently designed for selections contained a 40 nucleotide random as shown in **Figure 3.3.** I became the sole user of this pool and it has been solely used for robotic selections for the purpose of probing sequence space to ensure its integrity and to keep the lab-wide contamination to the minimum.

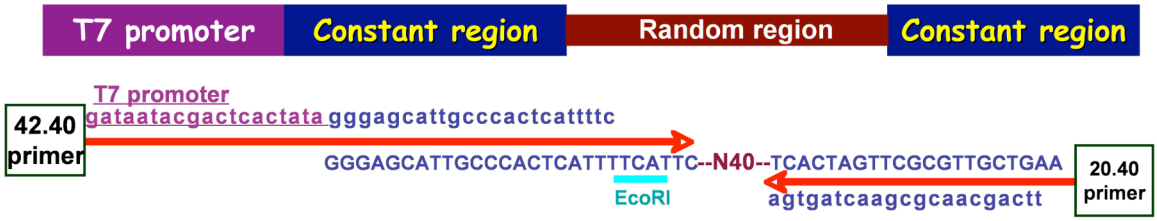


Figure 3.3. Graphical Representation of the N40 random pool. Capitalized residues represent bases transcribed into the RNA pool; lowercase bases represent priming

sequences. The 42.40 primer incorporates a T7 RNA polymerase promoter (gataatacgactcactata).

A single robotic run on the Biomek 2000 workstation performs six continuous rounds of aptamer selections in 18 hours. The progress of each robotic run is assessed through the amplification of DNA and formation of RNA for each round of selection. This is observed by running 10% of the collected reactions from each amplification step for all 18 rounds of selection. The reactions of each aliquot are analyzed on a denaturing 8% polyacrylamide gel to evaluate whether the RT-PCR reaction is successful in yielding dsDNA and whether the transcription (TXN) reactions produced RNA. This is an important step since robotic selections are continuous for 6 straight rounds, and there is no means by which the experimenter can control each step to make sure that every round is producing the nucleic acids needed for subsequent selections. For each pool, a total of 3 robotic selections are conducted to acquire data equivalent to 18 round of selection.

GLOBAL SELECTION ANALYSIS

All 8 N40 random pools underwent 18 rounds of selection against lysozyme with each pool's selection progress analyzed on an 8% polyacrylamide gel. **Figure 3.4** shows the PAGE gel analyses of the dsDNA and RNA products generated for each round of selection from each selected pool.

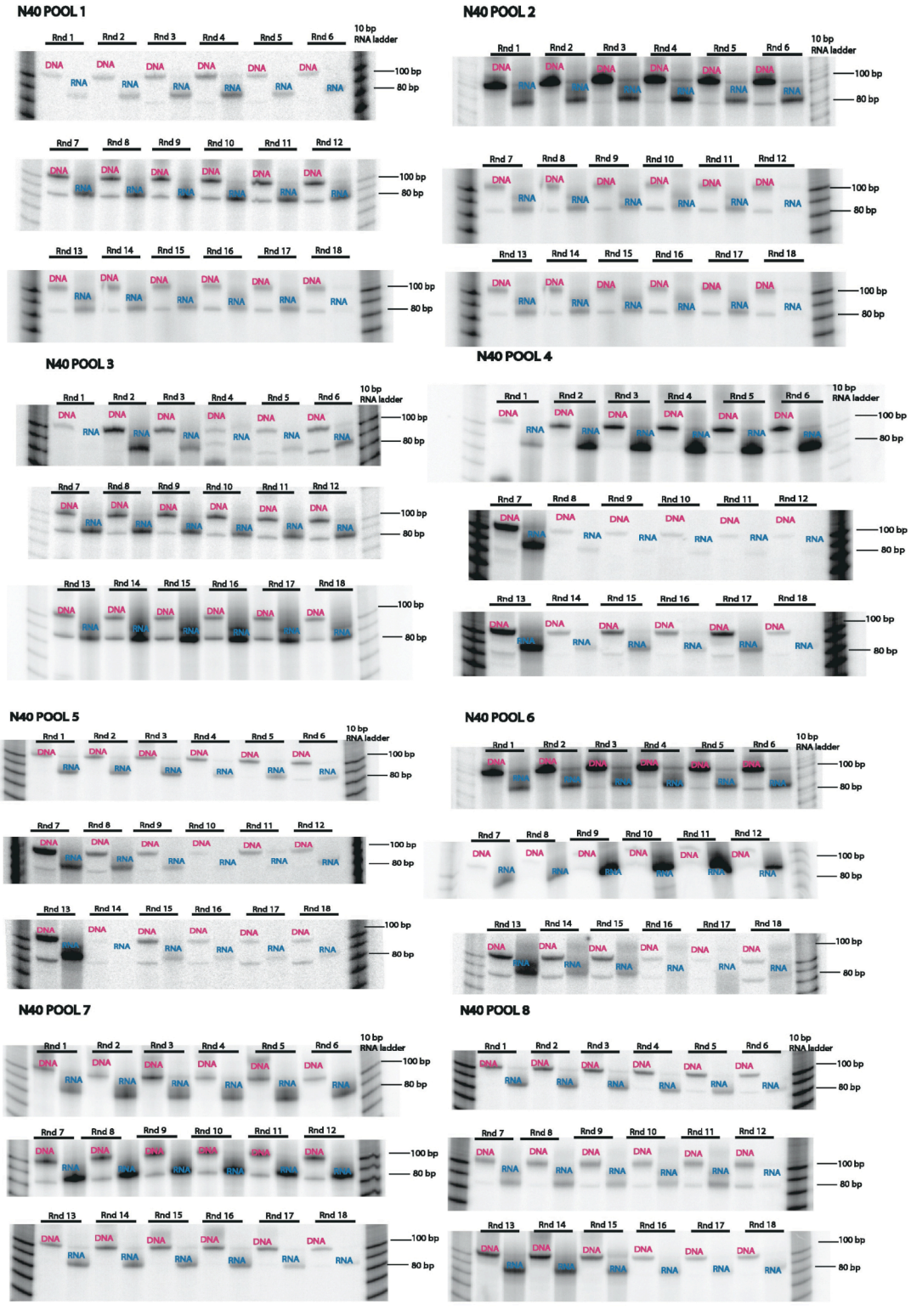

N40 POOL 1
Rnd 1
Rnd 2
Rnd 3
Rnd 4
Rnd 5
Rnd 6
10 bp RNA ladder
DNA
RNA
100 bp
80 bp
Rnd 7
Rnd 8
Rnd 9
Rnd 10
Rnd 11
Rnd 12
Rnd 13
Rnd 14
Rnd 15
Rnd 16
Rnd 17
Rnd 18
N40 POOL 2
N40 POOL 3
N40 POOL 4
N40 POOL 5
N40 POOL 6
N40 POOL 7
N40 POOL 8

Figure 3.4. Products Generated During the *in vitro* Selections of Pools 1-8. The figure shows the products that are isolated during the 18 round of selection against lysozyme. Each round of selection has one lane containing radiolabeled DNA from RT/PCR reactions, and another lane containing RNA transcribed products. The longer RT/PCR lane shows a product of 107 bases for the DNA template, while the shorter band in the transcription lane contains a product of 87 bases. A 10bp radiolabeled RNA ladder flanks the product lanes.

POOL 1 ANALYSIS

The next step that I carried out was to analyze the affinity of the selected pool to the protein target through a nitrocellulose-binding assay. Not every round was assayed for binding, except for round 0 (unselected RNA), round 6, round 12, and round 18. For each assay carried out, the main control utilized is clone 1, which previously isolated and highly characterized aptamers isolated from the N30 pool against lysozyme [8]. Binding assays involve the incubation of radiolabeled aptamers from each round against the target. The reaction is allowed to take place at room temperature and then passed through a filter for nitrocellulose and nylon as depicted in **Figure 3.5.**

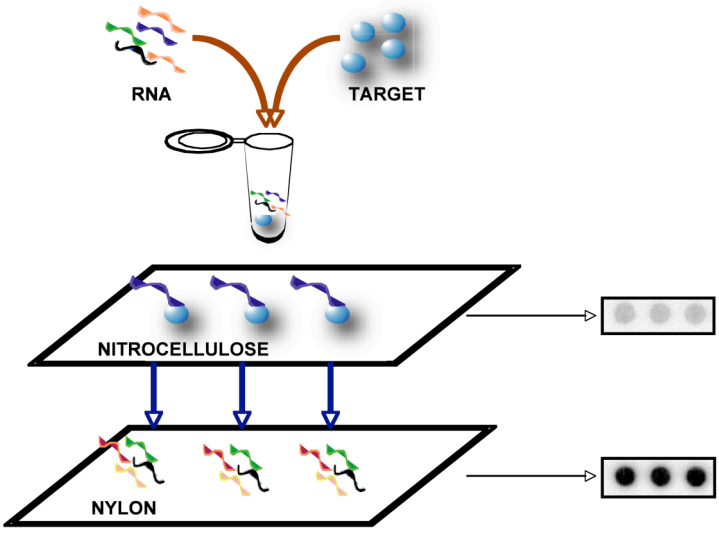


Figure 3.5. Nitrocellulose Binding Assays. Radiolabeled RNAs are incubated with the target of interest in selection buffer. After the incubation, the reaction is passed through two layers of filter, one nitrocellulose and the other nylon. RNAs that are bound to the target will remain in the top later of nitrocellulose along with the protein while the non-binders will flow through and remain in the nylon membrane.

It is important to use an aptamer control with a known behavior because previous data showed that nitrocellulose membranes vary from badge to badge leading to variations in the results. Using a highly characterized aptamer clone, we can standardize the numerical data from the binding assay reactions. To further ensure the fidelity of the data, all nitrocellulose-binding assays are carried out in triplicates and at least twice on different occasions.

Figure 3.6A illustrates the affinity trend of Pool 1 for free Lysozyme at different stages of the selections. Isolated RNAs were assayed against free Biotin to ensure that there was no nonspecific binding. The rationale for this step was to ensure that the winning species isolated were specific to the target and not to the biotin that has been conjugated to the protein target prior to the start of the selections. Pool affinity to the protein does not become obvious until selection round 12, which is expected of general selections. However, going from round 12 to round 18, there was a drop in binding affinity to the enzyme. There seemed to be a turnover of species going from round 12 to 18. Looking at the individual sequences isolated from each tested round, at round 12; there was an accumulation of sequence families. These families completely disappeared at round 18. There is a strong evidence of the loss of important binders.

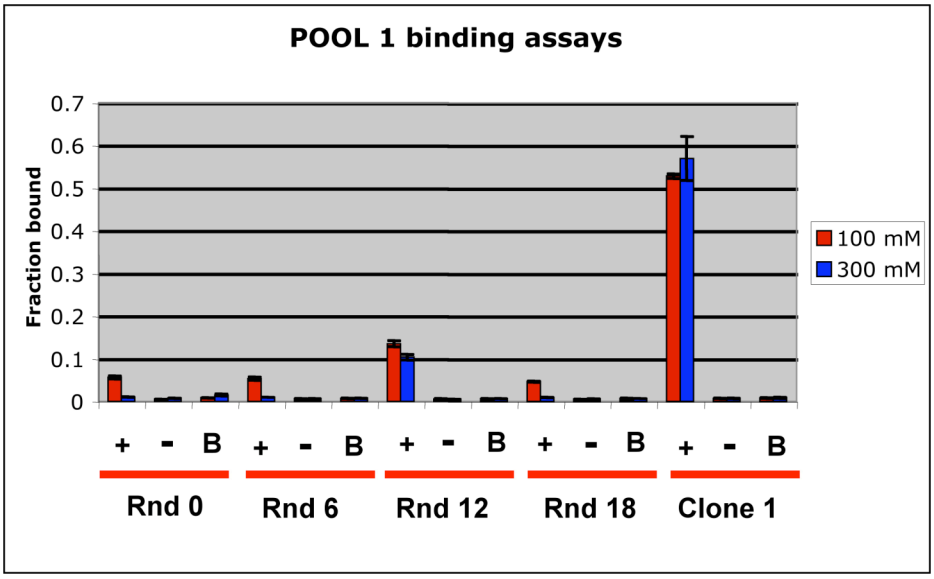


Figure 3.6A. Affinity Trends of Pool 1 for Lysozyme. RNA from different rounds were assayed and incubated with protein Lysozyme (+), incubated with only buffer (-), or with pure Biotin (B). Two different wash buffers were analyzed. Binding assays washed with a buffer identical to the selection buffer conditions is shown in red bars, and a high salt buffer wash containing 3X more salt (Sodium Chloride NaCl) than the selection buffer is shown in blue bars.

Looking at the isolated sequences from Pool 1 selections, the binding trend was translated into the sequences isolated (**Figure 3.6B).** From the binding data, it appears as though the pool at round 12 is slowly showing an improved binding, however, binding drops to almost background once it reaches round 18. At round 12 a predominant family was forming (A-3) among others, however, these binders disappeared at round 18. Not only did some of the sequences disappear, round 18 seems became more complex.

Figure 3.6B. Pool 1 Sequence Clones from Rounds 6, 12, and 18. N40 pool sequences

isolated from Pool 1 selection. The primer binding regions are colored maroon. Random regions are in black. Random regions that are highlighted signify identical sequences. Many sequences that were becoming dominant within round 12 of the isolated sequences dropped out at round 18. The binding of the aptamers also dropped from round 12 to 18.

A likely explanation for this observation is there may be a depletion of the good binders during the selection from round 12 to round 18. *In vitro* selection experiments are dependent on the amplification of functional species, and structural stability will discriminate between amplicons based on this trait rather than its ability to bind to the target. Stable RNA secondary structures that makes it difficult for the sequence to reach attain the conformation making it harder for polymerases to reach [12, 13]. An example of this is evidenced during *in vitro* selection of an RNA ligase. In this selection, RNA species with inactive conformations were isolated potentially because replication was more effective with conformational flexibility [14]. In another example, a selection for oligonucleotide targets for the T4 DNA ligase yielded mismatched species to the template because perfectly paired substrates were not as efficient during the amplification step [15].

POOL 2 ANALYSIS

Affinity trends for that of Pool 2 shows a gradual increase in the fraction of nucleic acids bound going from round 0 to round 12. No significant increases in binding affinity were further observed at round 18 (**Figure 3.7A)**.

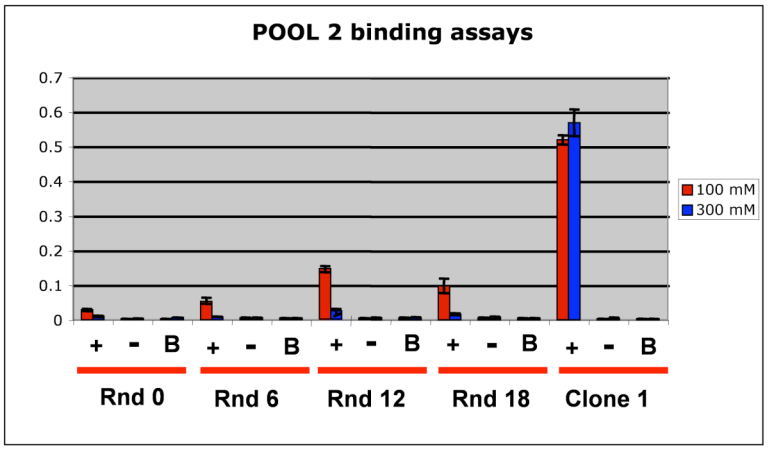


Figure 3.7A. Affinity Trends of Pool 2 for Lysozyme. RNA from different rounds were assayed and incubated with protein Lysozyme (+), incubated with only buffer (-), or with pure Biotin (B). Two different wash buffers were analyzed. Binding assays washed with a buffer identical to the selection buffer conditions is shown in red bars, and a high salt buffer wash containing 3X more NaCl than the selection buffer is shown in blue bars.

By looking at the sequences isolated from the assayed rounds (**Figure 3.7B**), after 18 rounds of selection, the pool was still highly complex without any predominant winning species. It could be inferred that these sequences were not evolving towards a region of better functionality. Traditionally during manual *in vitro* selections, one can increase the stringency so that the better binders and winnow down the pool size.

Round 6 sequences

```
>C1   GGGAGCATTGCCCACTCATTTTCATTC-CGTTCATACACTAAACTAGGTCAATCAGTCACGACGCTCG-GTCACTAGTTCGCGTTGCTG
>D1   GGGAGCATTGCCCACTCATTTTCATTC-TACAAGTAGGACAACTCCCTGAAAGACAAGCTTAGAGCAA-GTCACTAGTTCGCGTTGCTG
>E1   GGGAGCATTGCCCACTCATTTTCATTC-GGTAAATTCAGGAGAGAGATTCGTTTGACACAACGCCAAGGGTCACTAGTTCGCGTTGCTG
>F1   GGGAGCATTGCCCACTCATTTTCATTC-GTTTACTCAATCAGCCCGAAAGTGCGATACACACCGGGAT-GTCACTAGTTCGCGTTGCTG
>G1   GGGAGCATTGCCCACTCATTTTCATTC-GACGCGAGTTTTCTGTAATTGTTGGACGGATCGTCCACCG-GTCACTAGTTCGCGTTGCTG
>H1   GGGAGCATTGCCCACTCATTTTCATTC-AATTCATTTAGCTTGCAAGCATTTCACCACAACAGCCACA-GTCACTAGTTCGCGTTGCTG
>A2   GGGAGCATTGCCCACTCATTTTCATTC-GAAGCCTGTCAATCTGCAATCCGCCTTGCGGCTACAGTA--GTCACTAGTTCGCGTTGCTG
>B2   GGGAGCATTGCCCACTCATTTTCATTC-CCATGGAAGCCTGCAGGAAGTCCAGTCATAAACGAACTAC-GTCACTAGTTCGCGTTGCTG
>C2   GGGAGCATTGCCCACTCATTTTCATTC-GACTACGGCATTAACGCGGGCATTCCAGGGTTTACCTCGA-GTCACTAGTTCGCGTTGCTG
>E2   GGGAGCATTGCCCACTCATTTTCATTC-CTGCCCAATTCAGGGCTCTGAGAGTCGTGCGGTGAGTTAG-GTCACTAGTTCGCGTTGCTG
>F2   GGGAGCATTGCCCACTCATTTTCATTC-GGTCCTTTAACCCGCATGAGATATTCTCTGCCGCACATGG-GTCACTAGTTCGCGTTGCTG
>G2   GGGAGCATTGCCCACTCATTTTCATTC-CCGATATGGGACGCCTTCATCCATGTTTACCGTAACCAAG-GTCACTAGTTCGCGTTGCTG
>B3   GGGAGCATTGCCCACTCATTTTCATTC-CATTTTTACTAGATTGCTCCTAGGAAATTTGCAGCGTGTT-GTCACTAGTTCGCGTTGCTG
>C3   GGGAGCATTGCCCACTCATTTTCATTC-ACAGTGAGAATATTGCAAATATCATATTCGTGGGAACGGA-GTCACTAGTTCGCGTTGCTG
>D3   GGGAGCATTGCCCACTCATTTTCATTC-AGCATGTCATACTGATAACATACTCGCCAAAGAAGTAGGC-GTCACTAGTTCGCGTTGCTG
>E3   GGGAGCATTGCCCACTCATTTTCATTC-GTATCAAAGTAAAACGATCTTTACGTGGTGCCCTCCCTAT-GTCACTAGTTCGCGTTGCTG
>F3   GGGAGCATTGCCCACTCATTTTCATTC-CAATGTAGAAGAAATTCTCAGGGCATCTATAAGCGGTACC-GTCACTAGTTCGCGTTGCTG
>G3   GGGAGCATTGCCCACTCATTTTCATTC-ACGGGCTTTAAGAACCATTGGGATCATCAGTAAAGTTCAG-GTCACTAGTTCGCGTTGCTG
>H3   GGGAGCATTGCCCACTCATTTTCATTC-AACCAATTCGCTTGTTTATGCTCGCGTCTAATGTTCCTGG-GTCACTAGTTCGCGTTGCTG
>A4   GGGAGCATTGCCCACTCATTTTCATTC-CGAGAATCCGCTGTCAGAGAGGAACTGCATTCTTCTAGGG-GTCACTAGTTCGCGTTGCTG
>B4   GGGAGCATTGCCCACTCATTTTCATTC-GCCTTACGTGCGGCGCTCACTGCAATACGCGTGGATACGT-GTCACTAGTTCGCGTTGCTG
>C4   GGGAGCATTGCCCACTCATTTTCATTC-GAGCGGTCAACGCCGGTATGCGCCCGTGGAGGAAAGGGAC-GTCACTAGTTCGCGTTGCTG
>D4   GGGAGCATTGCCCACTCATTTTCATTC-GTTCATAAAAAGGCATAGACTGCTGTAATGCCAGCTTTAG-GTCACTAGTTCGCGTTGCTG
>E4   GGGAGCATTGCCCACTCATTTTCATTC-AGTTGGACTGAATTCGGTTCACGTCGTCTCGGACCACTAG-GTCACTAGTTCGCGTTGCTG
>F4   GGGAGCATTGCCCACTCATTTTCATTC-ATAGATACAACTCATATGCGCCAAATTCAGACTCGCGTCA-GTCACTAGTTCGCGTTGCTG
>G4   GGGAGCATTGCCCACTCATTTTCATTC-TAGTGTTCAATATCCTCCAGTTATATTCTCGAATGTCCGA-GTCACTAGTTCGCGTTGCTG
>H4   GGGAGCATTGCCCACTCATTTTCATTC-AGACCTCGACGAATTTAGTGGCTGAACCGTACCTCGAAAA-GTCACTAGTTCGCGTTGCTG
>A5   GGGAGCATTGCCCACTCATTTTCATTC-ATAGCGGCATAACCGCGAGTCCCTATTCACAATTTAAGAA-GTCACTAGTTCGCGTTGCTG
>B5   GGGAGCATTGCCCACTCATTTTCATTC-AGAGTAAGGTAAGAGCCACAGGTTCGTGCGGATTATCGTT-GTCACTAGTTCGCGTTGCTG
>D5   GGGAGCATTGCCCACTCATTTTCATTC-CCAAGCGAATCACCTATCAACCCTCGCACTGACTCCAAGA-GTCACTAGTTCGCGTTGCTG
>A9   GGGAGCATTGCCCACTCATTTTCATTC-TTCTACAACCTGTTAGATTAGGCGGTCCATTTTATCAATG-GTCACTAGTTCGCGTTGCTG
>B9   GGGAGCATTGCCCACTCATTTTCATTC-TCAGGTGAAGACGTCTAGACTCATAGCTCACAAGCTTATG-GTCACTAGTTCGCGTTGCTG
>C9   GGGAGCATTGCCCACTCATTTTCATTC-GGTAAAAGTTCTTAACGGCTTAATGCTGGGGGCGTGGTGT-GTCACTAGTTCGCGTTGCTG
>D9   GGGAGCATTGCCCACTCATTTTCATTC-ACCCAAAGCGGAAACGACACCTCCCATACAGGACCTGCCC-GTCACTAGTTCGCGTTGCTG
>E9   GGGAGCATTGCCCACTCATTTTCATTC-AGTCATCGTACCTGATGATTTGGACGTACCAACCTTCAAT-GTCACTAGTTCGCGTTGCTG
>F9   GGGAGCATTGCCCACTCATTTTCATTC-TTTCACTATAATAGCCGAATTTAATTCGAGCACTGTTTGT-GTCACTAGTTCGCGTTGCTG
```

Round 12 sequences

```
>A2   GGGAGCATTGCCCACTCATTTTCATTC-TGACAAGTGCAACCGTACACCGGTTTATGCTGGTAGAATA-GTCACTAGTTCGCGTTGCTG
>B4   GGGAGCATTGCCCACTCATTTTCATTC-TGACAAGTGCAACCGTACACCGGTTTATGCTGGTAGAATA-GTCACTAGTTCGCGTTGCTG

>G3   GGGAGCATTGCCCACTCATTTTCATTC-GCTCGTCGATTAGGGTCATAGCAATCATACTTTAATTAAG-GTCACTAGTTCGCGTTGCTG
>H3   GGGAGCATTGCCCACTCATTTTCATTC-GCTCGTCGATTAGGGTCATAGCAATCATACTTTAATTAAG-GTCACTAGTTCGCGTTGCTG
>D3   GGGAGCATTGCCCACTCATTTTCATTC-GACGTTTATTAACGGAATAGCCTACTGGTGCAACTACTGC-GTCACTAGTTCGCGTTGCTG
>F4   GGGAGCATTGCCCACTCATTTTCATTC-TTGTTAGGGTACTTTCTGTCAACGTGACATAGCTTTATTA-GTCACTAGTTCGCGTTGCTG
>B2   GGGAGCATTGCCCACTCATTTTCATTC-TATGTCCACAAACGACACTTCACACGATTTAAATTCCGCT-GTCACTAGTTCGCGTTGCTG
>C3   GGGAGCATTGCCCACTCATTTTCATTC-ACCGCACTCCAGAGGCTACGGGTTAAATTTCCTGACATGG-GTCACTAGTTCGCGTTGCTG

>F3   GGGAGCATTGCCCACTCATTTTCATTC-TAATTGAGTACAATTATTACATCGGCTATCTACCATAACT-GTCACTAGTTCG-GTTGCTG
>F2   GGGAGCATTGCCCACTCATTTTCATTC-GATGGGCCGACATCTATCCAGCAGCGCTATCGGAACATCC-GTCACTAGTTCGCGTTGCTG
>D2   GGGAGCATTGCCCACTCATTTTCATTC-CGGCGCTTTATGACTAGTATACGTCGGAGTATCTTAGTTA-GTCACTAGTTCGCGTTGCTG
>G1   GGGAGCATTGCCCACTCATTTTCATTC-AAAAACCTAAATTAGATTGCCTCCAAGCCATTAAGCGCGT-GTCACTAGTTCGCGTTGCTG
>H1   GGGAGCATTGCCCACTCATTTTCATTC-CCAGGTCTATTGTATACTACTCACTTACCATGTGGTTCGC-GTCACTAGTTCGCGTTGCTG
>G4   GGGAGCATTGCCCACTCATTTTCATTC-CAAATTACGTAATACTCGGCATTTCAAAATCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>C4   GGGAGCATTGCCCACTCATTTTCAT-C-GCGTGTTGCACAATCCCAAACGCCGTCGAATGAGACCAAC-GTCACTAGTTCGCGTTGCTG
>A4   GGGAGCATTGCCCACTCATTTTCATTC-ATCTGGTCGGAACATCATGGAATAATGCAGAGTCAGTCAT-GTCACTAGTTCGCGTTGCTG
>B1   GGGAGCATTGCCCACTCATTTTCATTC-CAGATTCTGACCCTTCTTCACGAGACACATAAGTAGCG---GTCACTAGTTCGCGTTGCTG
>G2   GGGAGCATTGCCCACTCATTTTCATTC-TGATACTAAACTAGTCAGTCGGAATTTCGCTGCAGGCCAG-GTCACTAGTTCGCGTTGCTG
>D4   GGGAGCATTGCCCACTCATTTTCATTC-TAACATATAGGGCAAAGCTCAAAACTCCTATTATATCGTTAGTCACTAGTT-GCGTTGCTG
>C2   GGGAGCATTGCCCACTCATTTTCATTC-CACCGGGCGGATGCTTGTGACTACACACCTAACTTTAAGG-GTCACTAGTTCGCGTTGCTG
>E2   GGGAGCATTGCCCACTCATTTTCATTC-CCTAAGGGTCATATTTGCTAGTACCTCTATCTGAGGCATA-GTCACTAGTTCGCGTTGCTG
>E1   GGGAGCATTGCCCACTCATTTTCATTC-ACGAAGTCCTAGGAAGAATAGGACGATCATGCAACTTGGTAGTCACTAGTTCGCGTTGCTG
>C1   GGGAGCATTGCCCACTCATTTTCATTC-TACCGGTTACGCATTAGTATAAAATTTGCGGCAGCAGA---GTCACTAGTTCGCGTTGCTG
>H4   GGGAGCATTGCCCACTCATTTTCATTC-GATTGCTCGAAACTTTTAGACTTAAATCAAGCAGGCGCAC-GTCACTAGTTCGCGTTGCTG
>A3   GGGAGCATTGCCCACTCATTTTCATTC-TAGCAGTCCCATTTTTGGGATAAGTACAGAAGAAGCTACA-GTCACTAGTTCGCGTTGCTG
>F1   GGGAGCATTGCCCACTCATTTTCATTC-TGCCAAGAATGGGAGACTTGGGTAACATTCGCGTCGTCCG-GTCACTAGTTCGCGTTGCTG
>E3   GGGAGCATTGCCCACTCATTTTCATTC-CACGCAACTGTAGGGCGTCTAACGTGTGTTGGAGCTGTCA-GTCACTAGTTCGCGTTGCTA
```

Round 18 sequences

```
>E9   GGGAGCATTGCCCACTCATTTTCATTC-GGAAGTGTAGGAGAACAGTAAGCTTAGAAACAATGGCCGGGGTCACTAGTTCGCGTTGCTG
>F10  GGGAGCATTGCCCACTCATTTTCATTC-GGAGGACAAAGTGGAACAGACTGTACAACAAGGCCGGC---GTCACTAGTTCGCGTTGCTG
>C10  GGGAGCATTGCCCACTCATTTTCATTC-GGCTTCGTACGGCGGGGTCCATAGGAAGATATCTCAGTG--GTCACTAGTTCGCGTTGCTG
>A7   GGGAGCATTGCCCACTCATTTTCATTC-CCCCCGACACTACCCCCGCAGCTACAACCAGGTCCCGGTC-GTCACTAGTTCGCGTTGCTG
>H8   GGGAGCATTGCCCACTCATTTTCATTC-CCCCCGCCGAACAGCTGCAGAAAGATAGAGCTCCTCCGCC-GTCACTAGTTCGCGTTGCTG
>C11  GGGAGCATTGCCCACTCATTTTCATTC-CCCCCTCGCAGGATGTAGAACTACAAAAACTCGGACCGCC-GTCACTAGTTCGCGTTGCTG
>C7   GGGAGCATTGCCCACTCATTTTCAT-C-CCCCCGCATAGGTACTAATACGAAAGCAAGATCGACCCGCC-GTCACTAGTTCGCGTTGCTG
>H11  GGGAGCATTGCCCACTCATTTTCATTC-CCCTAGGTGTGCTATACCCCGGAAAGAACTACGACCCCCC-GTCACTAGTTCGCGTTGCT-
>G9   GGGAGCATTGCCCACTCATTTTCATTC-CCCTCAACCGTGGCATAGGTAGGCTCAATGGACGGACGTG-GTCACTAGTTCGCGTTGCTG
>B9   GGGAGCATTGCCCACTCATTTTCATTC-CCTCCCGTGAAGAGTACGAACTAACAAACCTACCCCTGGC-GTCACTAGTTCGCGTTGCTG
>G12  GGGAGCATTGCCCACTCATTTTCATTC-CCGAACTGAGAACAGTCCGTAACGACAACCCCCACAGCCC-GTCACTAGTTCGCGTTGCTG
>H9   GGGAGCATTGCCCACTCATTTTCATTC-GTCAAAGCAAAGAGCTGACAGATAAGAACCAACCCCGGCC-GTCACTAGTTCGCGTTGCTG
>D9   GGGAGCATTGCCCTTTCATTTTCATTC-CCGGAATACAGAGAAGTAGTAACAAAGATTTGACCCGGCC-GTCACTAGTTCGCGTTGCTG
>B7   GGGAGCATTGCCCACTCATTTTCATTC-CCGTACACGCACCAACGACCAGCAGAGATACAACCCGGCC-GTCACTAGTTCGCGTTGCTG
>C8   GGGAGCATTGCCCACTCATTTTCATTT-CTTGTCTGACGCCGGCCAGTAGTCCGTCATGCACCCCGTGCGTCACTAGTTCGCGTTGCTG
>G10  GGGAGCATTGCCCACTCATTTTCATTC-CACCCTCCGAACTATACAAAACAAGACCGACCCGTCCTGC-GTCACTAGTTCGCGTTGCTG
>E11  GGGAGCATTGCCCACTCATTTTCATTC-AAGGATAGCGGGAGTAATCGTCGAGCCATGTTGTACGGGG-GTCACTAGTTCGCGTTGCTG
>E12  GGGAGCATTGCCCACTCATTTTCATTC-AGGCGGAGGACAAGACAGCATATGGGGACCCCCGGGTATC-GTCACTAGTTCGCGTTGCTG
>B11  GGGAGCATTGCCCACTCATTTTCATTC-ACAACAGAAACCCCCGATCCGAACCATTACCCCGACCCG--GTCACTAGTTCGCGTTGCTG
>H10  GGGAGCATTGCCCACTCATTTTCATTC-AGACCGAACAAGAACGCACATAATACCAAACCCCCCCGGAC-GTCACTAGTTCGCGTTGCTG
>D12  GGGAGCATTGCCCACTCATTTTCATTC-AGCAGCCCACCTCCGATCCCTCATACCTAACCCGCCGGTC-GTCACTAGTTCGCGTTGCTG
>E10  GGGAGCATTGCCCACTCATTTTCATTC-ACACCGAAGAACAGGATCATAGTACAAGCACACCGGGGCC-GTCACTAAGCCGAATTCCAG
>H12  GGGAGCATTGCCCACTCATTTTCATTC-ATCCCCAAAGGAACGGAGGGAAGAAGTCGATACAGGCCGG-GTCACTAGTTCGCGTTGCTG
>F9   GGGAGCATTGCCCACTCATTTTCATTC-ATGAGACCCCCTGCATTCCAACGAATGGAACCCCCCTGC--GTCACTAGTTCGCGTTGCT-
>H7   GGGAGCATTGCCCACTCATTTTCATTC-AATGGGGACGGAGAACAGAATCAACGATTGGCAGTTGGGG-GTCACTAGTTCGCGTTGCTG
>F7   GGGAGCATTGCCCACTCATTTTCATTC-AGGGAAGGCAGGGGAGAACCAAGAGGAACTTTGGCGCGTTGGTCACTAGTTCGCGTTGCTG
```

Figure 3.7B. Pool 2 Sequence Clones from Rounds 6, 12, and 18. N40 pool sequences isolated from Pool 1 selection. The primer binding regions are colored maroon. Random regions are in black. Random regions that are highlighted signify identical sequences. In this pool, there is no sign of pool enrichment in the different rounds of selection. In traditional selection methods, this result usually signifies that the selection has not gone to completion and additional manipulations, such as increasing in stringency, are needed to enrich the pool for better binders.

POOL 3 ANALYSIS

The binding data for Pool 3, more than two fold improvement in pool binding occured from round 12 to round 18 (**Figure 3.8A**). Evaluating the dynamics of pool 3 in terms of isolated clones from the selection (**Figure 3.8B),** it became clear that this increase in binding trend is expected. The round 12 pool was already highly enriched with one species of sequence populating half of the pool. At round 18, one predominant sequence clone overtook the entire pool. This is the general trend that is often observed in conventional aptamer selections.

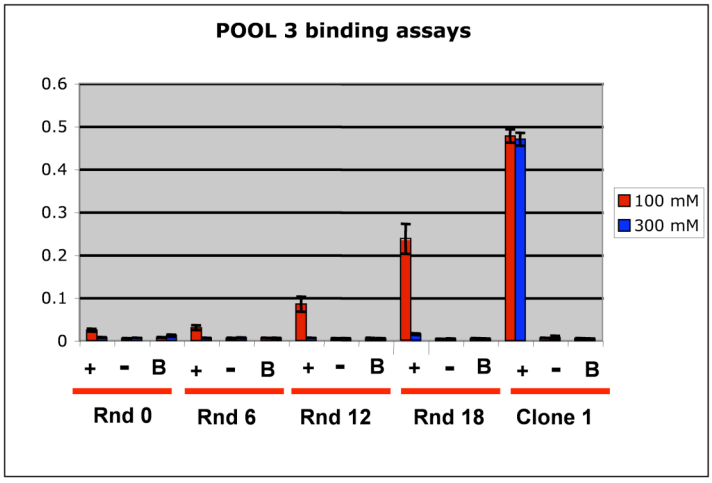


Figure 3.8A. Affinity Trends of Pool 3 for Lysozyme. RNA from different rounds were assayed and incubated with protein Lysozyme (+), incubated with only buffer (-), or with pure Biotin (B). Two different wash buffers were analyzed. Binding assays washed

with a buffer identical to the selection buffer conditions is shown in red bars, and a high salt buffer wash containing 3X more NaCl than the selection buffer is shown in blue bars.

Figure 3.8B. Pool 3 Sequence Clones from Rounds 6, 12, and 18. N40 pool sequences isolated from Pool 1 selection. The primer binding regions are colored maroon. Random

regions are in black. Random regions that are highlighted signify identical sequences. Underlined sequences show common motifs among the aptamer clones

POOL 4 ANALYSIS

Looking at the pools affinity towards the target Lysozyme, Pool 4 behaved in the same fashion as Pool 2 as shown **Figure 3.9A.** The pool did not seem to evolve at a great degree with binding only slightly better than the unselected pool, and no drastic change occurred from round 12 to round 18.

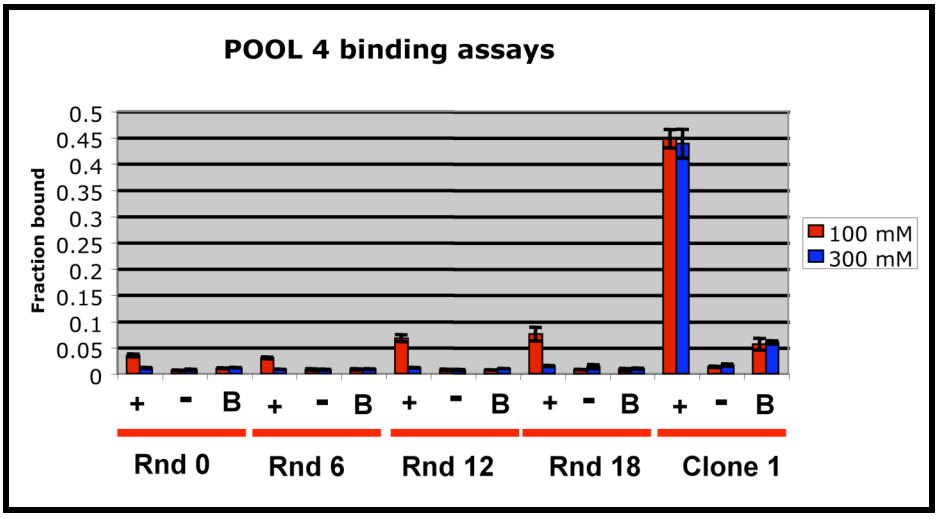


Figure 3.9A. Affinity Trends of Pool 4 for Lysozyme. RNA from different rounds were assayed and incubated with protein Lysozyme (+), incubated with only buffer (-), or with pure Biotin (B). Two different wash buffers were analyzed. Binding assays washed with a buffer identical to the selection buffer conditions is shown in red bars, and a high salt buffer wash containing 3X more NaCl than the selection buffer is shown in blue bars.

Round 6 sequences

```
>A1  GGGAGCATTGCCCACTCATTTTCATTC-ATCCTTGGTAAAACGTTCCGTTTCACTTGTAGGGT------GTCACTAGTTCGCGTTGCTG
>B1  GGGAGCATTGCCCACTCATTTTCATTG-GTTACATACAAGGACTGTCGAACTAGGGCGGGGAGAAC---GTCACTAGTTCGCGTTGCTG
>E1  GGGAGCATTGCCCACTCATTTTCATTC-TCGTACTCACACACAAACAATTAACTCCGCACATTCCCAG-GTCACTAGTTCGCGTT----
>H1  GGGAGCATTGCCCACTCATTTTCATTC-ATATTACCATGACAAAAACCCTGCGCCCGCTCCGTTAAAT-GTCACTAGTTCGCGTTGCT-
>G4  GGGAGCATTGCCCACTCATTTTCATTC-GGTTAACTCATCTCCCCGTTCCCTTTAACAACTACGTTGC-GTCACTAGTTCGCGT-GCTG
>H2  GGGAGCATTGCCCACTCATTTTCATTC-GAGTAACACTATCAGCCCTGAGGTTACGTCAATGGAGATC-GTCACTAGTTCGCGTTGCTG
>G2  GGGAGCATTGCCCACTCATTTTCATTC-CCAAAGTGACCACTCGTGTCAACTATTATGGGCACTCAAA-GTCACTAGTTCGCGTTGCTG
>G1  GGGAGCATTGCCCACTCATTTTCATTC-GCACGAACGCATTAGTTTTTACTACGCTCGATGGTGAATT-GTCACTAGTTCGCGTTGCTG
>H4  GGGAGCATTGCCCACTCATTTTCATTC-GACTGCCGAACGCGAGTCGAATATAACCCAGGGTTCAAAT-GTCACTAGTTCGCGT-GCTG
>H3  GGGAGCATTGCCCACTCATTTTCATTC-AGTAACTGAAAAGAGTACCCAGGCTTAGTCTAATTTTTGG-GTCACTAGTTCGCGTTGCTG
>F4  GGGAGCATTGCCCACTCATTTTCATTC-CCTGGCCGTGGAGACGGTCTAACTAAAGTAGCTTCTGATT-GTCACTAGTTCGCGTTGCTG
>A4  GGGAGCATTGCCCACTCATTTTCATTC-TCGTACTCACACACAAACAATTAACTCCGCACATTCCCAG-GTCACTAGTTCGCGTTTCTG
>E2  GGGAGCATTGCCCACTCATTTTCATTC-TAGCAGACGGCTCGGCAGGTGGAAGTTGGCCTTCTTCCTC-GTCACTAGTTCGCGTTGCTG
>C1  GGGAGCATTGCCCACTCATTTTCATTT-AGTGGAGAAAGTCACCCCCCGTCTTACGAATTATACCCC--GTCACTAGTTCGCGTTGCTG
>F1  GGGAGCATTGCCCACTCATTTTCATTC-ATGGTAGGGCTTCCCACTCCGTGCATGAAGCGCTGGAGGA-GTCACTAGTTCGCGTTGCTA
>C4  GGGAGCATTGCCCACTCATTTTCATTC-TTTCGTCTGGGTCGGTGACAAGCATCCAGCGGTCATGTAC-GTCACTAGTTCGCGTTGCTG
>C2  GGGAGCATTGCCCACTCATTTTCATTC-GGTTAACTCATCTCCCCGTTCCCTTTAACAACTACGTTGC-GTCACTAGTTCGCGT-GCTG
>B3  GGGAGCATTGCCCACTCATTTTCATTC-GCACCATGTTTCCGATCAGCGATGAGTGGCGATACCAAGG-GTCACTAGTTCGCGTTGCTG
>A3  GGGAGCATTGCCCACTCATTTTCATTC-AGGTTTCACTAAAGTGCAAAGCGTCAGTGAACTCCCATAA-GTCACTAGTTCGCGTTGCTA
>B4  GGGAGCATTGCCCACTCATTTTCATTC-GCATGAGCACCAAAGGGATTGCCATCTAAGCCCTCAACTG-GTCACTAGTTCGCGTTGCTG
>A2  GGGAGCATTGCCCACTCATTTTCATTC-TTTCGTCTGGGTCGGTGACAAGCATCCAGCGGTCATGTAC-GTCACTAGTTCGCGTTGCTG
>E4  GGGAGCATTGCCCACTCATTTTCATTC-TAGCAGACGGCTCGGCAGGTGGAAGTTGGCCTTCTTCCTC-GTCACTAGTTCGCGTTGCTG
>F2  GGGAGCATTGCCCACTCATTTTCATTC-ATCGTCGGCTAATGTGCCGACTAACATGCACGTAACGCAC-GTCACTAGTTCGCGTTGCTA
>B1  GGGAGCATTGCCCACTCATTTTCATTC-CCGGAGACATTGTCATCCCCGGCGGCTCCCCACATCTACA-GTCACTAGTTCGCGT-G--G
>B2  GGGAGCATTGCCCACTCATTTTCATTC-ATACCTAAGCCGCCAGTAGTATAATTTGAGTCATCTATTC-GTCACTAGTTCGCGTTGCTG
>D2  GGGAGCATTGCCCACTCATTTTCATTC-TGGCGACATCCACTGGCAACTTTGTAACCTAATGAACCTG-GTCACTAGTTCGCGTTGCTG
```

Round 12 sequences

```
>E6  GGGAGCATTGCCCACTCATTTTCATTC-ACTCCTTCCCTTATACCGTTGCCCCAATCTTTGCGCGTTAG-TCACTAGTTCGCGTTGCTG
>E8  GGGAGCATTGCCCACTCATTTTCATTC-ACTCCTTCCCTTATACCGTTGCCCCAATCTTTGCGCGTTAG-TCACTAGTTCGCGTTGCTG

>G7  GGGAGCATTGCCCACTCATTTTCATTC-TATTATGGATACAGAACTACGCAAAGGGAGTTAAGGGCTAG-TCACTAGTTCGCGTTACTG
>H7  GGGAGCATTGCCCACTCATTTTCATTC-TATTATGGATACAGAACTACGCAAAGGGAGTTAAGGGCTAG-TCACTAGTTCGCGTTACTG

>A7  GGGAGCATTGCCCACTCATTTTCATTC-GAGTGACACACGGCTAGCCCTTTGATTACGTCATACTAAAG-TCACTAGTTCGCGTTGCTG
>F6  GGGAGCATTGCCCACTCATTTTCATTC-AATCTTCAAACTATGACTCAACCACACATACAGTTGCCGTG-TCACTAGTTCGCGTTGCTG
>F5  GGGAGCATTGCCCACTCATTTTCATTC-TTTCGTGACAGGGCCCTTCCTATGTGTGTTTCCGAATACAG-TCACTAGTTCGCGTTGCTG
>D5  GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCGCTTCGGACG-TCACTAGTTCGCGTTGCTG
>H5  GGGAGCATTGCCCACTCATTTTCATTC-TATAGATGCGCTTTTACTAGGCAGGATGGGTTATAAAGGGG-TCACTAGTTCGCGTTGCTA
>D7  GGGAGCATTGCCCACTCATTTTCATTC-ACTGCAGATCCACGGGAAAATTGTAGAAAACGATAAAAGGG-TCACTAGTTCGCGTTGCTG
>D6  GGGAGCATTGCCCACTCATTTTCATTC-AGGAGCAAAATGATTCCCCACCGGTACTACGAAATTGCTAG-TCACTAGTTCGCGTTGCTA
>A5  GGGAGCATTGCCCACTCATTTTCATTC-GTACGGAGCTACCCAATAGCCGTTTGAATGCGTAAGCCGAG-TCACTAGTTCGCGTTGCTG
>G8  GGGAGCATTGCCCACTCATTTTCATTC-CGTAAAGTCCGATACTCTGATCGGACTGATATTGCCCCCGG-TCACTAGTTCGCGTTGCTG
>E5  GGGAGCATTGCCCACTCATTTTCATTC-AGCATGACATGTTGCACGCCAAGACGTTGCCTACTGAATAG-TCACTAGTTCGCGTTGCTG
>B5  GGGAGCATTGCCCACTCATTTTCATTC-TATCGTCTACATCATGAGCACTTAGTAAGTTGTAGTATCTG-TCACTAGTTCGCGTTGCTA
>B8  GGGAGCATTGCCCACTCATTTTCATTC-CCGATCTTGACTTATTTCACAACCCAAGGTTACGACTCTAG-TCACTAGTTCGCGTTACTG
>H6  GGGAGCATTGCCCACTCATTTTCATTC-CTCCCCTCGGGCCAGGCATAAAACTACAATGCAACTTGCGG-TCACTAGTTCGCGTTGCTG
>H8  GGGAGCATTGCCCACTCATTTTCATTC-GCATCTTATCCACTGGCAACCGAATTAAGATTCTTGCATGG-TCACTAGTTCGCGTTGCTG
>C5  GGGAGCATTGCCCACTCATTTTCATTC-GATCCGGCACAGCTAAGCACTTGAGAGGCATGGGCTGGCTG-TCACTAGTTCGCGTTACTA
```

Round 18 sequences

```
>C7  GGGAGCATTGCCCACTCATTTTCATTC-TCCGAGCATGGATCTACCGTGGACTACCATCGTTATCCTC-TCACTAGTTCGCGTTGCTG
>D6  GGGAGCATTGCCCACTCATTTTCATTC-TCCGAGCGGAGTCTGCTTGATTCTTTACCAAATCTATTAT-TCACTAGTTCGCGTTGCTG

>G5  GGGAGCATTGCCCACTCATTTTCATTC-AAATAACACCTGATTCAGTTTTAATCTTGTAATCGTGGT--TCACTAGTTCGCGTTGCTG
>E6  GGGAGCATTGCCCACTCATTTTCATTC-AAATAACACCTGATTCAGTTTTAATCTTGTAATCGTGGT--TCACTAGTTCGCGTTGCTG

>F8  GGGAGCATTGCCCACTCATTTTCATTC-TGTCCGGACTAACGCCAGTACGAGATTTTTCCGAGCCAGG-TCACTAGTTCGCGTTGCTA
>A6  GGGAGCATTGCCCACTCATTTTCATTC-AGCTACAGTTGCTGTGGAAATTTTCCGATGTGTGCGTCCA-TCACTAGTTCGCGTTGCTA

>E7  GGGAGCATTGCCCACTCATTTTCATTC-ATAAAATAAAATTGTAATCACGGTGCTCGAACTATGGCGTGTCACTAGTTCGCGTTGCTG
>H6  GGGAGCATTGCCCACTCATTTTCATTC-TTAGCACCAGCTAAAACACATAAAATATCCCTTGTGGCTG-TCACTAGTTCGCGTTGCTG

>D5  GGGAGCATTGCCCACTCATTTTCATTC-AGCGGCGCACTTGAGCGCAAGATCAGCAGGACTTCCGAGT-TCACTAGTTCGCGTTGCTA
>F7  GGGAGCATTGCCCACTCATTTTCATTC-CCGATCAGCGGCAGTTGGTTTCGCCGTGCCGTTCACAGA--TCACTAGTTCGCGTTGCTG
>A5  GGGAGCATTGCCCACTCATTTTCATTC-AGGGATTGCCCCTTCTCAGAGCGGTGCGCTCCTTTGGAGG-TCACTAGTTCGCGTTGCTG
>A8  GGGAGCATTGCCCACTCATTTTCATTC-TGCATCTCATCAGATGCTTGATGAGAAGCGCCTGTTCATC-TCACTAGTTCGCGTTGCTG
>G8  GGGAGCATTGCCCACTCATTTTCATTC-AATTTGTGTAAGTTAATCTTAGCTCCGGTGTAAGTCTGAG-TCACTAGTTCGCGTTGCTG
>H7  GGGAGCATTGCCCACTCATTTTCATTC-TCCTAACAAACTGTCCGATATTAACGGAACACCTATCCGA-TCACTAGTTCGCGTTACTA
>G6  GGGAGCATTGCCCACTCATTTTCATTC-ACGCAAACTCCAAGTCTCCGCAACACAGGATACCCCTTAC-TCACTAGTTCGCGTTGCTA
>C8  GGGAGCATTGCCCACTCATTTTCATTC-CGAACCAAGTTAAAATCTCGCACCCAGAGTGGAGATACC--TCACTAGTTCGCGTTGCTG
>E8  GGGAGCATTGCCCACTCATTTTCATTC-AATGTAGGTCCACATGTGCAAACTATAGTCAGTCAAAGGT-TCACTAGTTCGCGTTGCTG
>H8  GGGAGCATTGCCCACTCATTTTCATTC-TACCGCGGAATAGAGTGTCCAACGGCTTCCGTTGGGGCTG-TCACTAGTTCGCGTTGCTG
>C6  GGGAGCATTGCCCACTCATTTTCATTC-TAAGCTGGTATGCTTAGAATTCATGCCTTCTTGTAACGCC-TCACTAGTTCGCGTTACTG
>G7  GGGAGCATTGCCCACTCATTTTCATTC-GGCGTCCGGTTAGTGATTGGTAATGCCGGACGGTGGTGTA-TCACTAGTTCGCGTTGCTG
>B6  GGGAGCATTGCCCACTCATTTTCATTC-GGTTGGATCTAATGACGCCAGGAGTGACGGGCACTCGATG-TCACTAGTTCGCGTTGCTG
>A7  GGGAGCATTGCCCACTCATTTTCATTC-TGATTGCTAGTTACTAGCAGAATAAACCTTGTCTGATGC--TCACTAGTTCGCGTTACTA
```

Figure 3.9B. Pool 4 Sequence Clones from Rounds 6, 12, and 18. N40 pool sequences isolated from Pool 1 selection. The primer binding regions are colored maroon. Random regions are in black. Random regions that are highlighted signify identical sequences. Underlined sequences show common motifs among the aptamer clones.

POOL 5 ANALYSIS

The binding assay for Pool 5 shows that the isolated pools' affinity to Lysozyme is much higher (fractions higher than 0.15 as opposed to 0.10) as shown in **Figure 3.10A.** The improvement in binding followed a gradual trend. Moreover, the affinity was improved to about 0.25, a significant improvement. This trend signals the reduction in pool complexity to a few strong binders [8, 9, 16, 17].

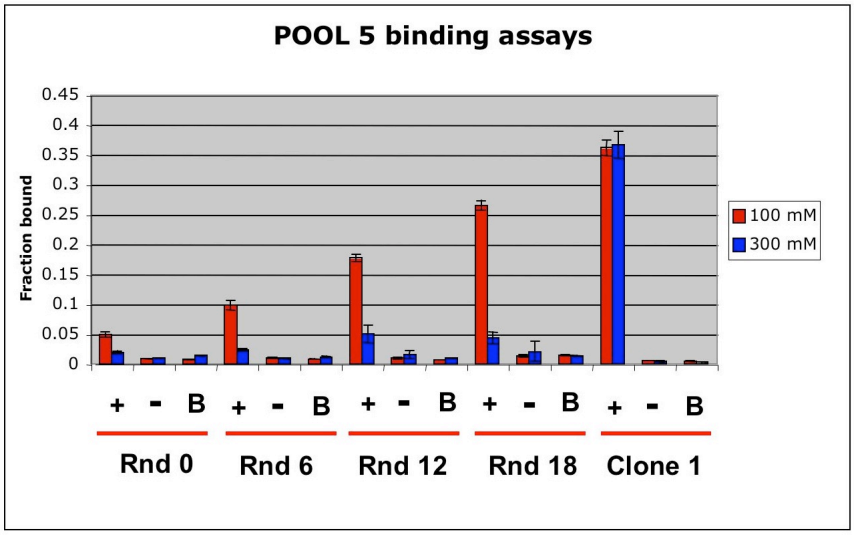


Figure 3.10A. Affinity Trends of Pool 5 for Lysozyme. RNA from different rounds were assayed and incubated with protein Lysozyme (+), incubated with only buffer (-), or with pure Biotin (B). Two different wash buffers were analyzed. Binding assays washed with a buffer identical to the selection buffer conditions is shown in red bars, and a high salt buffer wash containing 3X more NaCl than the selection buffer is shown in blue bars.

From the sequences isolated, there was no predominant clone that was found in the pool, however, a distinctive GC rich motif towards the 3' end of the random region was observed across all round 18 sequences. (**Figure 3.10B)**

```
Round 6 sequences

>H12  GGGAGCATTGCCCACTCATTTTCATTC-ACGCTCAATGGGTCAGGGGACAAAGGGAGTCGACACTA---GTCACTAGTTCGCGTTGCTG
>H11  GGGAGCATTGCCCACTCATTTTCATTC-CCTCATAGGCCCTCCAACTATAACCTGAGCTAGCACTGGC-GTCACTAGTTCGCGTTGCTG
>D10  GGGAGCATTGCCCACTCATTTTCATTC-GGGAACACCGATGGTAGAATACCTCTGCGACACGGCCCTA-GTCACTAGTTCGCGTTGCTG
>H10  GGGAGCATTGCCCACTCATTTTCATTC-ACCGTGCACATGTTACGTACAAATATCCGAACCCGGGGCC-GTCACTAGTTCGCGTTGCTG
>G9   GGGAGCATTGCCCACTCATTTTCATTC-TCACGCAAAAAGGCACCTGCATTACATAACCGCCCCCTGT-GTCACTAG-TCGCGTTGCTG
>G11  GGGAGCATTGCCCACTCATTTTCATTC-GACGTCGACATTATGAAATAGTCCAATCACAAAGAGGCCG-GTCACTAGTTCGCGTTGCTG
>G10  GGGAGCATTGCCCACTCATTTTCATTC-TTCTTCACGCCGAGGCCGGAGGGAGGTGTAACGTCGGT---GTCACTAGTTCGCGTTGCTG
>G12  GGGAGCATTGCCCACTCATTTTCATTC-AATTGGCAGGGGAGATCGTCTGTGGTCCGTCGGGGC-----GTCACTAGTTCGCGTTGCTG
>C10  GGGAGCATTGCCCACTCATTTTCATTC-AGAGTCCAGCCTATAATAACGAGCGGAACATCGTCGGTG--GTCACTAGTTCGCGTTGCTG
>D9   GGGAGCATTGCCCACTCATTTTCATTC-GTAGCTCTACGTGTAACGACTGGATAATTCCCCGCGGGGC-GTCACTAGTTCGCGTTGCTG
>E12  GGGAGCATTGCCCACTCATTTTCATTC-GTAGCTCTACGTGTAACGACTGGATAATTCCCCGCGGGGC-GTCACTAGTTCGCGTTGCTG
>D12  GGGAGCATTGCCCACTCATTTTCATTC-GTAGCTCTACGTGTAACGACTGGATAATTCCCCGCGGGGC-GTCACTAGTTCGCGTTGCTG
>A10  GGGAGCATTGCCCACTCATTTTCATTC-AGTCTACAAGGTCTACTCGAGCATCGGATCGACACCTCTG-GTCACTAGTTCGCGTTGCTG
>H9   GGGAGCATTGCCCACTCATTTTCATTC-GTGGACAAATAGAACAAAGTTTCTTACGTTATGTCGTGGG-GTCACTAGTTCGCGTTGCTG
>A11  GGGAGCATTGCCCACTCATTTTCATTC-AGGCACCAATAGCGCCCTGGGAAAGGCGGCCGACCGTGCC-GTCACTAGTTCGCGTTGCTG
>E10  GGGAGCATTGCCCACTCATTTTCATTC-CGCGGACCAGTGTACCGCCCCTAAGATACGCCTCGCATCC-GTCACTAGTTCGCGTTGCTG
>F10  GGGAGCATTGCCCACTCATTTTCATTC-TCGACGCTCTGAGACATGAATGTTCTAATCCCGGGTGCCC-GTCACTAGTTCGCGTTGCTG
>E9   GGGAGCATTGCCCACTCATTTTCATTC-CGCAACAATATGTAGCGAGACCACATCAGGGGCACATGGG-GTCACTAGTTCGCGTTGCTG
>B11  GGGAGCATTGCCCACTCATTTTCATTC-GTAAGCGCTGATTCAGTTAGCGCCAATACTAGATGGCCC--GTCACTAGTTCGCGTTGCTG
>B12  GGGAGCATTGCCCACTCATTTTCATTC-TGCCAACTACATCCGTCCCGACGACCAGCTAGGCCACCTT-GTCACTAGTTCGCGTTGCTG
>A9   GGGAGCATTGCCCACTCATTTTCATTC-GACCAAATGGACGACAGGGAGATGTAAAACACGGCCACTC-GTCACTAGTTCGCGTTGCTG
>A12  GGGAGCATTGCCCACTCATTTTCATTC-ACCGTGCACATGTTACGTACAAATATCCGAACCCGGGGCC-GTCACTAGTTCGCGTTGCTG
>C11  GGGAGCATTGCCCACTCATTTTCATTC-GCTCACCACGGCCATGATATCAAGAGGAGGCTCCGCGCGG-GTCACTAGTTCGCGTTGCTG
>E11  GGGAGCATTGCCCACTCATTTTCAT-C-GGGTCTCGGGCCAAATCCAAATGGCGATAAGAAAGGCTGC-GTCACTAGTTCGCGTTGCTG

Round 12 sequences

>F6   GGGAGCATTGCCCACTCATTTTCATTC-GCTTTCTATGGCCCCTGGTATCAGCAGACCCCGCCAGTGT--GTCACTAGTTCGCGTTGCTG
>A4   GGGAGCATTGCCCACTCATTTTCATTC-GCTTTCTATGGCCCCTGGTATCAGCAGACCCCGCCAGTGT--GTCACTAGTTCGCGTTGCTG

>A3   GGGAGCATTGCCCACTCATTTTCATTC-GACCCACAGTATCACTATCGCCAAACAGCAGAAGACCGCC--GTCACTAGTTCGCGTTGCTG
>F3   GGGAGCATTGCCCACTCATTTTCATTC-GACCCAGAGTAATGATCATCCCGGAACTGCCCCCCAGC----GTCACTAGTTCGCGTTGCTG

>G4   GGGAGCATTGCCCACTCATTTTCATTC-AAAGAAGGCGAGGAAACGAATCGAGAACATAGCGGCCGGG--GTCACTAGTTCGCGTTGCTG
>H6   GGGAGCATTGCCCACTCATTTTCATTC-AAAGGTCAACGAACAGAGAGTAGAACAAGAACCGCCGGGG--GTCACTAGTTCGCGTTGCTG
>B2   GGGAGCATTGCCCACTCATTTTCATTC-GTAGCGTCAGGGACAATAGGGAACATGTAAATGCCGGCGG--GTCACTAGTTCGCGTTGCTG
>E1   GGGAGCATTGCCCACTCATTTTCATTC-TAGCAACTCTCAAGAACTTTACGGCCTACATATCGGCGGG--GTCACTAGTTCGCGTTGCTG
>A5   GGGAGCATTGCCCACTCATTTTCATTC-GGAACGACCAGCTAAGCACCGCAATCTTACCA-CGCCGGG--GTCACTAGTTCGCGTTGCTG
>B4   GGGAGCATTGCCCACTCATTTTCATTC-CCCGAAAAGTTATGACCCCCAAACCTACTGAACTGGCGCC--GTCACTAGTTCGCGTTGCT-
>F2   GGGAGCATTGCCCACTCATTTTCATTC-GACCCCGCATAAGAACAAGCAGATCAAACGGTACCCGGCC--GTCACTAGTTCGCGTTGCTG
>D4   GGGAGCATTGCCCACTCATTTTCATTC-GACCCGCGCTAATAAGACAAAGGAATAAATTACGCCCGCC--GTCACTAGTTCGCGTTGCTG

>D2   GGGAGCATTGCCCACTCATTTTCATTC-CCCAACTCAAACCCCAGCGAATTTATAACGCCCCGGCATG--GTCACTAGTTCGCGTTGCTA
>E3   GGGAGCATTGCCCACTCATTTTCATTC-ACTAAGGTAACTCACAAGACGGACTAACATCCCCCCGGGC--GTCACTAGTTCGCGTTGCTG

>E2   GGGAGCATTGCCCACTCATTTTCATTC-CAGGCACGGGAATCACAACTACACAATTCCTCCTGGTGCG--GTCACTAGTTCGCGTTGCTG
>D1   GGGAGCATTGCCCACTCATTTTCATTC-CGCCGAAACGACGTTGCGACTACCAAGCCCGGGGCCCACC--GTCACTAGTTCGCGTTGCTG
>C1   GGGAGCATTGCCCACTCATTTTCATTC-CCCTCCCCACCAGAACCAAGAAACACTCAATTTGCCAGGCC-GTCACTAGTTCGCGTTGCTG
>C2   GGGAGCATTGCCCACTCATTTTCATTC-GGAAACAGTAAGAAGAGCAGAGATACCCACCCGGTCCGC---GTCACTAGTTCGCGTTGCTG
>F1   GGGAGCATTGCCCACTCATTTTCATTC-GAAAGCGGGAGTCATAACACTACTACGAGCATACGGGGCC--GTCACTAGTTCGCGTTGCTG
>H3   GGGAGCATTGCCCACTCATTTTCATTC-GGATGGACAAGAATATCGTTTGGCTAACAACCCGGGGGTG--GTCACTAGTTCGCGTTGCTG
>G3   GGGAGCATTGCCCACTCATTTTCATTC-GGATGGGGAAAAGGATGACTCACAAACAACAGATGGTGGC--GTCACTAGTTCGCGTTGGTG

Round 18 sequences

>F8   GGGAGCATTGCCCACTCATTTTCATTC-CCCCTGCGATGGAAGAGGAAACAGACATGAACTGGCCCGG-GTCACTAGTTCGCGTTGCTG
>C12  GGGAGCATTGCCCACTCATTTTCATTC-ACGAACAGAACCAGCCTCACGACCCCTCCAAACCCCGGCC-GTCACTAGTTCGCGTTGCTG
>G12  GGGAGCATTGCCCACTCATTTTCATTC-CGCGACCCTATTGAACACCCCGCCCTTGGAACCCCCGGCC-GTCACTAGTTCGCGTTGCTG
>E10  GGGAGCATTGCCCACTCATTTTCATTC-GACCCGTCGCAGACCCCATCAGAGCTACTGATCGCCCCCG-GTCACTAGTTCGCGTTGCTG
>H7   GGGAGCATTGCCCACTCATTTTCATTC-GACCCGAAGCCCAAAGAACCTAGAGAGCAACCACCCGGCC-GTCACTAG-TCGCGTTGCTG
>C11  GGGAGCATTGCCCACTCATTTTCATTC-GCGAAACCAACCAAGCCCCGTATGGAAGATAGACCCGGCC-GTCACTAGTTCGCGTTGCTG
>H12  GGGAGCATTGCCCACTCATTTTCATTC-ACGTAGCGAGAGAACGATAAGTACAGAGAACACGGCCCGG-GTCACTAGTTCGCGTTGCTG
>G11  GGGAGCATTGCCCACTCATTTTCATTC-GCTGAGAGGGTGCAGGTCTACTGAATGGGCTAGCCGTGGG-GTCACTAGTTCGCGTTGCTG
>G9   GGGAGCATTGCCCACTCATTTTCATTC-GAAGCGAAGATAAAGCAAGAAGCACTATCCAGGCCGGCC--GTCACTAGTTCGCGTTGCTG
>G7   GGGAGCATTGCCCACTCATTTTCATTC-GGTGGGAGGACTACGTCAAGCAACTACATACATGGGGCCG-GTCACTAGTTCGCGTTGCTG
>E11  GGGAGCATTGCCCACTCATTTTCATTC-GAACAGTAGAGAAAGCAGAAGACGTAAACGAAGCGCCGGG-GTCACTAGTTCGCGTTGCTG
>F12  GGGAGCATTGCCCACTCATTTTCATTC-AAGGACGGGGGCAGAACACGCATCAGAACTTAACGGCCGG-GTCACTAGTTCGCGTTGCTG
>E7   GGGAGCATTGCCCACTCATTTTCATTC-GCAGGGGGAAGCTTGTAGAACAATGGTACCGACGCCGGGG-GTCACTAGTTCGCGTTGCTG
>H9   GGGAGCATTGCCCACTCATTTTCATTC-ACAGAACCGACATGACCCCCGCAGCTACAAGACCGCCGCC-GTCACTAGTTCGCGTTGCTG
>D8   GGGAGCATTGCCCACTCATTTTCATTC-AGGGAAGGCCACAATAGACGAACACAAGTATCCGCGGGG--GTCACTAGTTCGCGTTGCTG
>E12  GGGAGCATTGCCCACTCATTTTCATTC-GAAACCAGAAAATATCCCAACGTCAATCAGACACGGCCGG-GTCACTAGTTCGCGTTGCTG
>B11  GGGAGCATTGCCCACTCATTTTCATTC-CCACAAAAAGGCAAAGAGTATCGAGAAGGTCACCCCCGGC-GTCACTAGTTCGCGTTGCTG
>A9   GGGAGCATTGCCCACTCATTTTCATTC-GCCAGCGTTTGCCACAGACACCACAACGACATCGGCCCGG-GTCACTAGTTCGCGTTGCTG
>A10  GGGAGCATTGCCCACTCATTTTCATTC-AGCACAGAGTACGTAAGGATAAGGAATACAACTGCCGGGG-GTCACTAGTTCGCGTTGCTG
>F11  GGGAGCATTGCCCACTCATTTTCATTC-CAGGAAGCTAGAAGGGAGGAACAAACAGCTTGAACCGCCC-GTCACTAGTTCGCGTTGCTG
>B9   GGGAGCATTGCCCACTCATTTTCATTC-GCAGGATGGGTAAGGCAAATTGGACAGTTGATTAGCGGGG-GTCACTAGTTCGCGTTGCTG
```

Figure 3.10B. Pool 5 Sequence Clones from Rounds 6, 12, and 18. N40 pool sequences isolated from Pool 1 selection. The primer binding regions are colored maroon.

Random regions are in black. Random regions that are highlighted signify identical sequences. Underlined sequences show common motifs among the aptamer clones.

POOL 6 ANALYSIS

This pool's binding trend showed a lot of parallel to that shown for Pool 3 (**Figure 3.11A).** The ability of the pool to bind to the target at round 18 was high. This can only be explained by the loss of pool complexity at this stage with the presence of species winners (**Figure 3.11B)**.

.

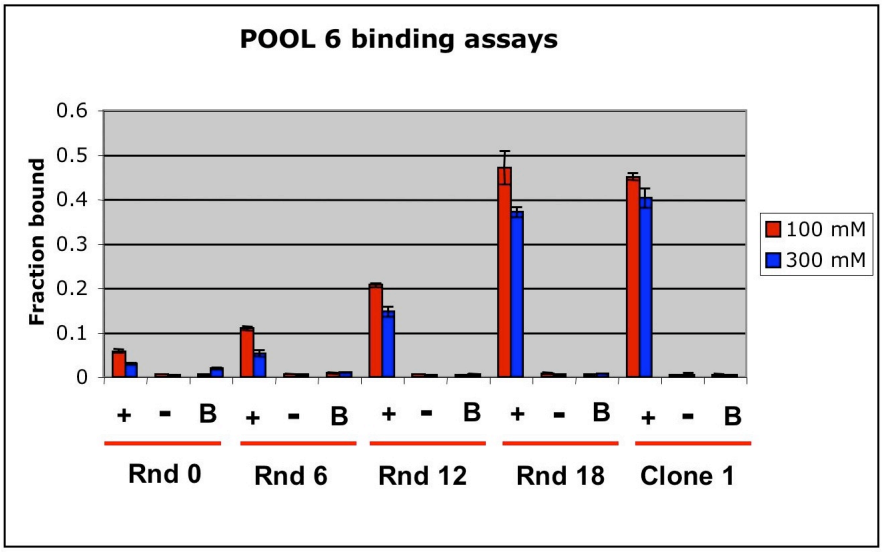


Figure 3.11A. Affinity Trends of Pool 2 for Lysozyme. RNA from different rounds were assayed and incubated with protein Lysozyme (+), incubated with only buffer (-), or with pure Biotin (B). Two different wash buffers were analyzed. Binding assays washed with a buffer identical to the selection buffer conditions is shown in red bars, and a high salt buffer wash containing 3X more NaCl than the selection buffer is shown in blue bars.

Round 6 sequences

```
>D1 GGGAGCATTGCCCACTCATTTTCATTC-TCACGCAACAACACTCGACAGCGAACCGACATTCCCCAGG-GT-ACTAGGTCCGGTTGGTG
>D2 GGGAGCATTGCCCACTCATTTTCATTC-GATGGTCTGCAACAAGTGTCAGGTAGTGTGAAGGGGCTGG-GTCACTAGTTCGCGTTGCTG
>A4 GGGAGCATTGCCCACTCATTTTCATTC-GGAGCAGAGTGGTGTCAAAGTAGGCTTAAGCTGCGTGGTG-GTCACTAGTTCGCGTTGCTG
>C3 GGGAGCATTGCCCACTCATTTTCATTC-CCGCCAGTCACTCCCTCTGACCGCCTAGCCCCCGCATCCC-GTCACTAGTTCGCGTTGCTG
>B4 GGGAGCATTGCCCACTCATTTTCATTC-AGCCCTCGAGTAATCAGGATCCGACGTCGTCTCGCTTATC-GTCACTAGTTCGCGTTGCTG
>D4 GGGAGCATTGCCCACTCATTTTCATTC-CAAGCGAAAGTCAACAGGAGAGAGATGGTACGAACCGATC-GTCACTAGTTCGCGTTGCTG
>E4 GGGAGCATTGCCCACTCATTTTCATTC-GCCGGAACAACCGCTAGGCCCAGTTACCAACCGCCCTGTG-GTCACTAGTTCGCGTTGCTG
>F1 GGGAGCATTGCCCACTCATTTTCATTC-GCACCGTACCGATCCACATCGGGGTAGTTCTTGATGGCGTGGTCAATAGTTCGCGTTAAGC
>F2 GGGAGCATTGCCCACTCATTTTCATTC-GTCCGCTCAAACCGCAAATCAAGTATGAGCCCCCGCCGCC-GTCACTAGTTCGCGTTGCTG
>E2 GGGAGCATTGCCCACTCATTTTCATTC-GGGAACGTTATAGAACATTTTGGAAGGAGTGCCGCACGGG-GTCACTAGTTCGCGTTGCTG
>H2 GGGAGCATTGCCCACTCATTTTCATTC-ACGAGGCGAAACGGTAGCTAGTCTCGGTCCCTCCCCCGGC-GTCACTAGTTCGCGTTGCTG
>H1 GGGAGCATTGCCCACTCATTTTCATTC-AGACAGCAGCAATCCCCACGAATAAGAAGGCCTGAGCTTA-GTCACTAGTTCGCGTTGCTG
>B3 GGGAGCATTGCCCACTCATTTTCATTC-TAGCAAATTCCCGGCGGGGCTGATAGCATCTAGCGTTATG-GTCACTAGTTCGCGTTGCTG
>F3 GGGAGCATTGCCCACTCATTTTCATTC-CCATCACGGACACCACTGATCAGCAGGCCCCGACCCAACG--TCACTAGTTCGCGTTGCTG
>C2 GGGAGCATTGCCCACTCATTTTCATTC-CCCAGGCACAATACACGAAGTTCCGAGCGGCCGCCTTAAT-GTCACTAGTTCGCGTTGCTG
>A2 GGGAGCATTGCCCACTCATTTTCATTC-ACGGCTAGGATTGCGATGTGCACAACAGGAATGGACGGTG-GTCACTAGTTCGCGTTGCTG
>C1 GGGAGCATTGCCCACTCATTTTCATTC-CCCTCGACTAGAATCCGATTTTGCAACCCCGCCAGCTGTC-GTCACTAGTTCGCGTTGCTG
>E3 GGGAGCATTGCCCACTCATTTTCATTC-AGCCAGCAGACCGGTGTCCCGCATCTTCACGCACATGGCCCGTCACTAGTTCGCGTTGCTG
>H4 GGGAGCATTGCCCACTCATTTTCATTC-TCAGCCAATCGGGTACATACGGAGAAACACATCGCGCCTG-GTCACTAGTTCGCGTTGCTG
>D3 GGGAGCATTGCCCACTCATTTTCATTC-AAATGACACAACAGCTGTAGTCGGCGGGCTTATGAACAAC-GTCACTAGTTCGCGTTGCTG
```

Round 12 sequences

```
>H7 GGGAGCATTGCCCACTCATTTTCATTC-GCCACACCCGCCCAAGACCCTCGGATTTAGGACCTCGCCC-GTCACTAGTTCGCGG--CTA
>H6 GGGAGCATTGCCCACTCATTTTCATTC-TTGGCCCGGCGGCATAACTATACATTGAACCCCCACCCGC-GTCACTAGTTCGCGTTGCTG
>F8 GGGAGCATTGCCCACTCATTTTCATTC-GAACGATCCCACGTGAAACAGAAGTAGACTCTGACCGGCC-GTCACTAGTTCGCGTTGCTG
>C8 GGGAGCATTGCCCACTCATTTTCATTC-CCCGCACCTCACGTTAACAGTCCTCCTCTACAACGGCCCC-GTCACTAGTTCGCGTTGCTG
>E7 GGGAGCATTGCCCACTCATTTTCATTC-CGCCCCGCCACGCCGTAGCAGCTATACCTTACCCCGCCTG-GTCACTAGTTCGCGTT-CTG
>B7 GGGAGCATTGCCCACTCATTTTCATTC-CCGCCTGCCGAAAGCCCGCGCCCTTACACGTCTGACCGCC-GTCACTAGATCGCGTTGCTG
>C5 GGGAGCATTGCCCACTCATTTTCATTC-CCCCTGAGAGCGTAGATAGTACACAACCCCCCGACGGTGC-GTCACTAGTTCGCGTTGCTG
>C7 GGGAGCATTGCCCACTCATTTTCATTC-GACCCCGTACAGTAGACGCAAATAATTCCCCCCCACCGTC-GTCACTAGTTCGCGTTGCTG
>D5 GGGAGCATTGCCCACTCATTTTCATTC-AGGGACCCGGACCTTAAGTCGACCACCAATCGCCGCACAC-GTCACTAGTTCGCGTTGCTG
>D8 GGGAGCATTGCCCACTCATTTTCATTC-CCGACCAGAGGCTCAAACAGATATCCCCGGATTCCCCGC--GTCACTAGTTCGCGTTGCTG
>A6 GGGAGCATTGCCCACTCATTTTCATTC-CCCAGCAACCTCTTCCCCAGCTAATAACTCATTCCCCGGC-GTCACTAGTTCGCGTTGCTG
>H8 GGGAGCATTGCCCACTCATTTTCATTC-GGTGAAACTACGAAAACGATGTGCAATAAACAGCCGGGCC-GTCACTAGTTCGCGTTGCTG
>B6 GGGAGCATTGCCCACTCATTTTCATTC-CCACACGATGCCCTCCTCATGGACATTTGACAGGACCCCC-GTCACTAGTTCGCGTTGCTG
>A8 GGGAGCATTGCCCACTCATTTTCATTC-GCACTACGACTGAGTAGACAACTGATAGCATAGGCCGGCC-GTCACTAGTTCGCGTTGCTG
>H5 GGGAGCATTGCCCACTCATTTTCATTC-CCGCAGCCCCCCGACACAGCCACAATATTAACCACCCGGC-GTCACTAGTTCGCGTTGCTG
>B5 GGGAGCATTGCCCACTCATTTTCATTC-GAACCTGTCGTGTGCGTACCGAATAACCTGTAAGCCGGCC-GTCACTAGTTCGCGTTGCTG
>F7 GGGAGCATTGCCCACTCATTTTCATTC-CAGTTCTAGCGACCACCCCCAACCGAAATAACGCCAGGCC-GTCACTAGTTCGCGTTGCTG
>E6 GGGAGCATTGCCCACTCATTTTCATTC-CCGTGCACCCATACAGAGATCGAACTACTATCCGCCGGCC-GTCACTAGTTCGCGTTGCTG
>A7 GGGAGCATTGCCCACTCATTTTCATTC-AGTGTGAGGGCTAATATCATCAAGATCGAACTACGGGCGG-GTCACTAGTTCGCGTTGCTG
>G7 GGGAGCATTGCCCACTCATTTTCATTC-GACGAAGACGTCCCCCAGTACCCCCGACACGCGCCCTTG--GTCACTAGTTCGCGTTGCTG
```

Round 18 sequences

```
>E5 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>E6 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>E4 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>E2 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>C5 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>C2 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>B2 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>H6 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAGCATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>A6 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>H1 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG

>H2 GGGAGCATTGCCCACTCATTTTCATAC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGG---GTCACTAGTTCGCGTTGCTG
>G1 GGGAGCATTGCCCACTCATTTTCATTC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGGG--GTCACTAGTTCGCGTTGCTG
>D6 GGGAGCATTGCCCACTCATTTTCATTC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGGG--GTCACTAGTTCGCGTTGCTG
>A2 GGGAGCATTGCCCACTCATTTTCATTC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGG---GTCACTAGTTCGCGTTGCTG
>B3 GGGAGCATTGCCCACTCATTTTCATTC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGGG--GTCACTAGTTCGCGTTGCAG
>F1 GGGAGCATTGCCCACTCATTTTCATTC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGGG--GTCACTAGTTCGCGTTGCTG


>F2 GGGAGCATTGCCCACTCATTTTCATTC-GAGCAGCAAACTGGAAGGGATAAGATAAGAAGTCGCCGGG--GTCACTAGTTCGCGTTGCTG
>H3 GGGAGCATTGCCCACTCATTTTCATTC-GAGCAGCAAACTGGAAGGGATAAGATAAGAAGTCGCCGGG--GTCACTAGTTCGCGTTGCTG
>E1 GGGAGCATTGCCCACTCATTTTCATTC-GAGCAGCAAACTGGAAGGGATAAGATAAGAAGTCGCCGGG--GTCACTAGTTCGCGTTGCTG
>A3 GGGAGCATTGCCCACTCATTTTCGTTC-GAGCAGCAAACTGGAAGGGATAAGATAAGAAGTCGCCGGG--GTCACTAGTTCGCGTTGCTG

>E3 GGGAGCATTGCCCACTCATTTTCATTC-GACCCCCCGGATAATCGACCCGCAGAGCCTAGACCCCGCC--GTCACTAGTTCGCGTTGCTG
>B6 -GGAGCATTGCCCACTCATTTTCATTC-GACCCCCCGGATAATCGACCCGCAGAGCCTAGACCCCGCC--GTCACTAGTTCGCGTTGCTG
>G5 GGGAGCATTGCCCACTCATTTTCATTC-GACCCCCCGGATAACCGACCCGCAGAGCCTAGACCCCGCC--GTCACTAGTTCGCGTTGCTG

>F4 GGGAGCATTGCCCACTCATTTTCATTC-GAAGCGTCAGCAGTCACGGAATACATTGTCGGTCGGGGTG--GTCACTAGTTCGCGTTGCTG
>H4 GGGAGCATTGCCCACTCATTTTCATTC-GAAGCGTCAGCAGTCACGGAATACATTGTCGGTCGGGGTG--GTCACTAGTTCGCGTTGCTG

>G6 GGGAGCATTGCCCACTCATTTTCA--C-CGAACCATACACCTTGACCACCCCCCAAGAGAGCCCCGCC--GTCACTAGTTCGCGTTGCTG
>A4 GGGAGCATTGCCCACTCATTTTCA--C-CGAACCATACACCTTGACCACCCCCCAAGAGAGCCCCGCC--GTCACTAGTTCGCGTTGCTG

>G3 GGGAGCATTGCCCACTCATTTTCATTC-CCTCCACGAGTATAAGAACAAAAGACAGAAACCCCCGGGCCCGTCACTAGTTCGCGTTGCTG
>B1 -GGAGCATTGCCCACTCATTTTCATTC-CCTCCACGAGTATAAGAACAAAAGACAGAAACCCCCGGGCCCGTCACTAGTTCGCGTTGCTG

>D5 GGGAGCATTGCCCACTCATTTTCATTC-CCAACCCCGCTAAGAACAAGAAAGCCACGAGTCCCCGGCC--GTCACTAGTTCGCGTTGCTG
>G2 GGGAGCATTGCCCACTCATTTTCATTC-CCAACCCCGCTCAGAACAAGAAAGCCACGAGTCCCCGGCC--GTCACTAGTTCGCGTTGCTG

>C1 GGGAGCATTGCCCACTCATTTTCATTC-AACAGCGGATAGAACGTAGCTATAAGACAGTTTGGCGGGG--GTCACTAGTTCGCGTTGCTG
>D1 GGGAGCATTGCCCACTCATTTTCATTC-AACAGCGGATAGAACGTAGCTATAAGACAGTTTGGCGGGG--GTCACTAGTTCGCGTTGCTG
>C6 GGGAGCATTGCCCACTCATTTTCATTC-AACGGATAGAGCATAAAGTCGAAGCATCAACTTGGCGGG---GTCACTAGTTCGCGTTGCTG

>C3 GGGAGCATTGCCCACTCATTTTCATCC-GAACGATAACGAAGATACGCCCTAGATTTCCCCGCCGCC---GTCACTAGTTCGCGTTGCTG
>F3 GGGAGCATTGCCCACTCATTTTCATTC-GAATGAAGGCTAAACGACAAGGAAAGACATTTCGGCGGGG--GTCACTAGTTCGCGTTGCTG

>A5 GGGAGCATTGCCCACTCATTTTCATTC-GAAGAGGAGAGTAAACAGGGACGAACACCGTACTGGCCGG--GTCACTAGTTCGCGTTGCTG
>A1 -GGAGCATTGCCCACTCATTTTCATTC-CCGCGATACCTGCCCCAACGAACGAGACCACAGACCCGCC--GTCACTAGTTCGCGTTGCTA
>F5 GGGAGCATTGCCCACTCATTTTCATTC-GCAACGACAACATCGCTTGAGATACATACATGCCACGGGG--GTCACTAGTTCGCGTTGCTG
>H5 GGGAGCATTGCCCACTCATTTTCATTC-GCGCCAGAAATCGCCCGACCCAGCACCACTTTGGACCGCC--GTCACTAGTTCGCGTTGCTG
```

Figure 3.11B. Pool 6 Sequence Clones from Rounds 6, 12, and 18. N40 pool sequences isolated from Pool 1 selection. The primer binding regions are colored maroon. Random regions are in black. Random regions that are highlighted signify identical sequences. Underlined sequences show common motifs among the aptamer clones.

POOL 7 ANALYSIS

The binding for this pool shows similar trend as that of Pool 1-3, just at a much higher affinity (closer to 0.2 as opposed to 0.1) (**Figure 3.12A).** As it can be predicted now, the complexity of the pool will remain high across all analyzed rounds (**Figure 3.12B).**

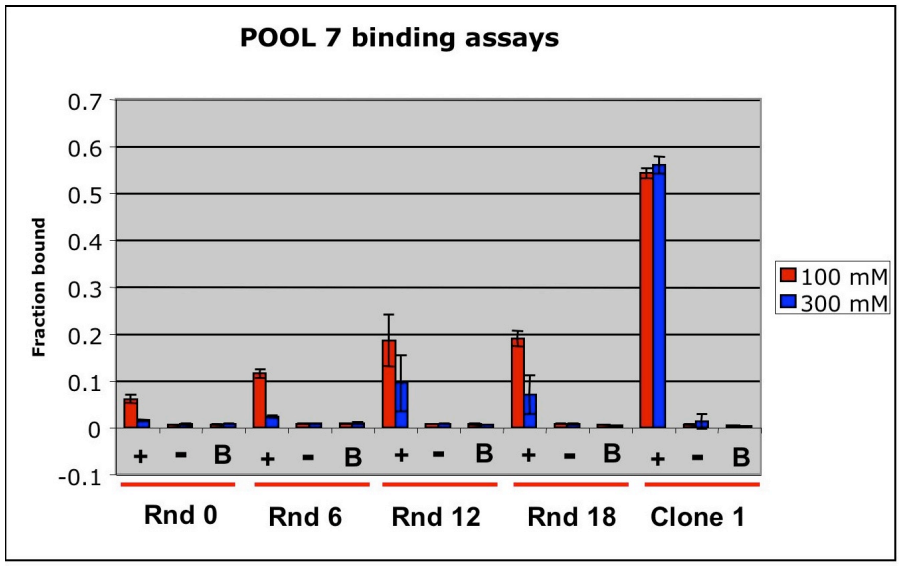


Figure 3.12A. Affinity Trends of Pool 7 for Lysozyme. RNA from different rounds were assayed and incubated with protein Lysozyme (+), incubated with only buffer (-), or with pure Biotin (B). Two different wash buffers were analyzed. Binding assays washed with a buffer identical to the selection buffer conditions is shown in red bars, and a high salt buffer wash containing 3X more NaCl than the selection buffer is shown in blue bars.

Round 6 sequences

```
>C2 GGGAGCATTGCCCACTCATTTTCATTC-ACGCTAGA-GTTAGCTTAGGATAGACTTCCTCTCGATGTAC-GTCACTAGTTCGCGTTGCTG
>D6 GGGAGCATTGCCCACTCATTTTCATTC-TCGCCC--CTCTGCTCCAGACGTGAGTTTGGTTCTCGGTCGGGTCACTAGTTCGCGTTGCTG
>D5 GGGAGCATTGCCCACTCATTTTCATTC-CAACAGAACTGTCATAACCCCCGTAATC---AACGTTGCTC-GTCACTAGATCGCGTTGCTG
>A4 GGGAGCATTGCCCACTCATTTTCATTC-TAGATCAATCGTACAGGT-CGACACGACGTTGCCGCGTCT-CGTCACTAGTTCGCGTTGCTG
>A2 GGGAGCATTGCCCACTCATTTTCATTC-AGCTGC--TCTAACTGACCTTAAGTCTTGGCAGTTTAACGAGGTCACTAGTTCGCGTTGCTG
>B1 -GGAGCATTGCCCACTCATTTTCATTC-ACAAGA--GGCACCGCGCGGCCACCATATCGATTACACAAGTGTCACTAGTTCGCGTTGCTG
>E4 GGGAGCATTGCCCACTCATTTTCATTC-GATCCACAGTTTCGTCAC--CGTCTTTCTTTGTATCGCTGGTGTCACTAGTTCGCGTTGCTG
>E3 GGGAGCATTGCCCACTCATTTTCATTC-TAACTCAACGTTAGCTTGCAGAGAAACTACCTCCAGTTTG--GTCACTAGTTCGCGTTGCTG
>H6 GGGAGCATTGCCCACTCATTTTCATTC-TTCCCA--CATCACCAGGTTCCACCGTGAGGAGCACAACACTGTCACTAGTTCGCGTTGCTG
>G3 GGGAGCATTGCCCACTCATTTTCATTC-TCTAATTGTGC--CTTTCCGTATCAATTCGTAGATATATACGGTCACTAGTTCGCGTTGCTG
>A1 GGGAGCATTGCCCACTCATTTTCATTC-TCGTAC--TCCCCTCAATATTGCACAAATGAACTCTCAGCAGGTCACTAGTTCGCGTTGCTG
>D4 GGGAGCATTGCCCACTCATTTTCATTC-TGGACC--CTAGTGAACAGACGTCGGAACATTATCTAGTATTGTCACTAGTTCGCGTTGCTG
>C5 GGGAGCATTGCCCACTCATTTTCATTC-GTTAAG--TATCCAGGACTCGGCAGTGGACGATCGGAGACCCGTCACTAGTTCGCGTTGCTG
>C4 GGGAGCATTGCCCACTCATTTTCATTC-CAACAGAACTGTCATAACCCCCGTAATC---AACGTTGCTC-GTCACTAGTTCGCGTTGCTG
>G6 GGGAGCATTGCCCACTCATTTTCATTC-ACGGGA--ACAAGCCAGGCGCCATCA-ATATCCTACATACGCGTCACTAGTTCGCGTTGCTG
>G5 GGGAGCATTGCCCACTCATTTTCATTC-GTTTCCGTTGT-GAAACTTTAGCCTAGACCGC-CTCCACCTAGTCACTAGTTCGCGTTGCTG
>B3 GGGAGCATTGCCCACTCATTTTCATTC-GAAATCTGTCCGTAAGTAGTTCTGGAGTTGTGTAGTGTGG--GTCACTAGTTCGCGTTGCTG
>H5 GGGAGCATTGCCCACTCATTTTCATTC-ATCCAT--CTGTACCTTGCCAGACTGAATTAGCACCTTCGCTGTCACTAGTTCGCGTTGCTG
>H3 GGGAGCATTGCCCACTCATTTTCATTC-AGGCACGCGAAGCCGGTCATACTCCTACACGT-CGC-CTCACGTCACTAGTTCGCGTTGCTG
>D2 GGGAGCATTGCCCACTCATTTTCATTC-AACAATCGATC--CTGCTCATAGAGCCAACTAGGCTTTGATCGTCACTAGTTCGCGTTGCTG
>C3 GGGAGCATTGCCCACTCATTTTCATTC-CATAGACATTCCTTAAACGTTATGGTCTAGACCAACCATG--GTCACTAGTTCGCGTTGCTG
>B4 GGGAGCATTGCCCACTCATTTTCATTC-GCGCTGTTCCACGAAATTGATCATACCTCA-AGCTTCGTTC-GTCACTAGTTCGCGTTGCTG
>G2 GGGAGCATTGCCCACTCATTTTCATTC-TCAATA--AGTCTCCTAAGTGACCGCTCGGAACCACATAACCGTCACTAGTTCGCGTTGCTG
>E6 GGGAGCATTGCCCACTCATTTTCATTC-GAAGTGGGAGT--AGTCAATAGCCCAGCACATACGGCACGGCGTCACTAGTTCGCGTTGCTG
>E5 GGGAGCATTGCCCACTCATTTTCATTC-A-AGCCGCAAGGTGCGCGATAGCGGACGTGTAGTGCAAATT-GTCACTAGTTCGCGTTGCTG
>A6 GGGAGCATTGCCCACTCATTTTCATTC-CACCAACAGCTTGCTACTGCTGGAGAGAATCGTGGGAGTT--GTCACTAGTTCGCGTTGCTG
```

Round 12 sequences

```
>B6  GGAGCATTGCCCACTCATTTTCATTC-CCCGCAAAAAACGCAGCATCCGTCTTATAGCATCCCCGGC----GTCACTAGTTCGCGTTGCTG
>F2  GGAGCATTGCCCACTCATTTTCATTC-CAAAAACTTAACGCCCTGTAGCCATGAATTAGCCCCCCGA----GTCACTAGTTCGCGTTGCTG
>F3  GGAGCATTGCCCACTCATTTTCATTC-CCGCTCCAACGTAGAAAAGCAAGAGTACGCCCTCGCCTAG----GTCACTAGTTCGCGTTGCTG
>C6  GGAGCATTGCCCACTCATTTTCATTC-CCTGAGAAATAGGGCACCACGGAAGAATCGTACAGCGGAGG---GTCACTAGTTCGCGTTGCTG
>C1  GGAGCATTGCCCACTCATTTTCATTC-CATGAGGTGGGTTACAACGGTCATACCACGAACCTGGCGG----GTCACTAGTTCGCGTTGCTG
>C4  GGAGCATTGCCCACTCATTTTCATTC-CGGAAAGACAGCACTCACAATCCAGTCACGACGTGGGGG-----GTCACTAGTTCGCGTTGCTG
>G2  GGAGCATTGCCCACTCATTTTCATTC-AAGCCACTGCCATGAACGATCATGAGACCATGATCACCGGCCGGGTCACTAGTTCGCGTTGCTG
>B3  GGAGCATTGCCCACTCATTTTCATTC-AACAGAGGAGGACGCGTTTCACCCTAGAACCCAACGCCCCCCACGTCACTAGTTCGCGTTGCTG
>G6  GGAGCATTGCCCACTCATTTTCATTC-ACACGAGAACAGCATTAATGGCTTATTGAGAGATGGCGGG----GTCACTAGTTCGCGTTGCTG
>D3  GGAGCATTGCCCACTCATTTTCATTC-ACACAACGAATCGAAAGCCATAAAACACTCATCTGCGGGG----GTCACTAGTTCGCGTTGCTG
>B1  GGAGCATTGCCCACTCATTTTCATTC-AGAAGAATTTCGGGGCGTATTTACACTCTGCCGGGGTGC-----GTCACTAGTTCGCGTTGCTG
>D4  GGAGCATTGCCCACTCATTTTCATTC-AGTAACCAATTCTCGCTGTCACACTATCGGGGCATGGTGG----GTCACTAGTTCGCGTTGCTG
>E4  GGAGCATTGCCCACTCATTTTCATTC-GCACGTGGCTGCCAAGAGGACCTGAGTAGCAATCAGGCAGCCTGGTCACTAGTTCGCGTTGCTG
>E1  GGAGCATTGCCCACTCATTTTCATTC-GCAGAATAGTTAGAGTAGCTCAGGACCGACCCCCT---------GTCACTAGTTCGCGTTGCTG
>F1  GGAGCATTGCCCACTCATTTTCATTC-GAACAGGAAGACCTCACACAGTACAATCCATGGATGGGCC----GTCACTAGTTCGCGTTGCTG
>D2  GGAGCATTGCCCACTCATTTTCATTC-GCCTGACGCTACGAAACGATAGAACGAACACATACGCCCCCGG-GTCACTAGTTCGCGTTGCTG
>D1  GGAGCATTGCCCACTCATTTTCATTC-GTCAATCAAGGGACGTTGGGATAAACTCTGGGGCTGGTAC----GTCACTAGTTCGCGTTGCTG
>E6  GGAGCATTGCCCACTCATTTTCATTC-GCGACCCCCGAACAATTTCGTAAGATATCTCTAGCCCGCC----GTCACTAGTTCGCGTTGCTG
>F5  GGAGCATTGCCCACTCATTTTCATTC-GGGCAGGACTGAAGGTGGGTTTCAGTTTGAGAGAGCGCTG----GTCACTAGTTCGCGTTGCTG
>E5  GGAGCATTGCCCACTCATTTTCATTC-GGGTAACACAGAATGGGCGGCCTAAAAGAAAACGACCGGC----GTCACTAGTTCGCGTTGCTG
```

Round 18 sequences

```
>D9  GGGAGCATTGCCCACTCATTTTCATTC-GCAGAATAAAGCCCCGTAATACGACCTTTGGACCCGCC---GTCACTAGTTCGCGTTGCTG
>F9  GGGAGCATTGCCCACTCATTTTCATTC-GAAGAGCACTGGGGATGAATCAGCATGTTTGTGCACCGGG-GTCACTAGTTCGCGTTGCTG
>G9  GGGAGCATTGCCCACTCATTTTCATTC-GACAAAGAATAACAAAGGCACAAGAGCACAGTATGGGGGG-GTCACTAGTTCGCGTTGCTG
>B10 GGGAGCATTGCCCACTCATTTTCATTC-TGCAAACACAGACCCCCCGGAAGGCGATCAAAACACCCCGCGTCACTAGTTCGCGTTGCTG
>C10 GGGAGCATTGCCCACTCATTTTCATTC-ACCAAGAAAAGAGGAGAAAACACAGGGAACTATCGGCG---GTCACTAGTTCGCGTTGCTG
>G10 GGGAGCATTGCCCACTCATTTTCATTC-AGGGAGCACAATGGAGTCTACAAGCGTAAGTCTGGCCGGG-GTCACTAGTTCGCGTTGCTG
>F12 GGGAGCATTGCCCACTCATTTTCATTC-GACCCCCAACAACTACGAATATGTACAACGACCGGCCCTT-GTCACTAGTTCGCGTTGCTG
>A10 GGGAGCATTGCCCACTCATTTTCAT-C-CCCCGCCGAAAGGTAGAGAAAACACGACCAGAGACCGCCC-GTCACTAGTTCGCGTTGCTG
>E10 GGGAGCATTGCCCACTCATTTTCATTC-CGAGGCAAGGGAAGTTAGAGGTCAAGAGATTAGCGCTGGG-GTCACTAGTTCGCGTTGCTG
>D10 GGGAGCATTGCCCACTCATTTTCATTC-GAAGGACGAGCAGGGTATACACCAAATGGAACACGGCGGG-GTCACTAGTTCGCGTTGCTG
>E11 GGGAGCATTGCCCACTCATTTTCATTC-AGGGCAGGAAGAAACTGGTAGGATGCACTTGACGCCGGGG-GTCACTAGTTCGCGTTGCTG
>F10 GGGAGCATTGCCCACTCATTTTCATTC-GAGTTGGGGAAGGGCAGTCACATAGAATAAGGTCGGGCC--GTCACTAGTTCGCGTTGCTG
>H11 GGGAGCATTGCCCACTCATTTTCATTC-GGAGGGGCTAAGGACACAAGAGCAAACACTTGGCATGGGG-GTCACTAGTTCGCGTTGCTG
>H10 GGGAGCATTGCCCACTCATTTTCATTC-GGGGGGACAACGTAAACATGAGGCAAGGATCGCAGTTCAC-GTCACTAGTTCGCGTTGCTG
>E12 GGGAGCATTGCCCACTCATTTTCATTC-CCGTAGGCAGCTAGAACCAAAAGTCGAGAGTACCCGGCCC-GTCACTAGTTCGCGTTGCTG
```

Figure 3.12B. Pool 7 Sequence Clones from Rounds 6, 12, and 18. N40 pool sequences isolated from Pool 1 selection. The primer binding regions are colored maroon. Random regions are in black. Random regions that are highlighted signify identical sequences..

POOL 8 ANALYSIS

The binding trend, albeit with higher fraction bound, has a very similar look to that of Pool 4. The binding peaks at about 0.30 at round 12 and gets significantly reduced again at round 18 (**Figure 3.13A).** The binding trend strongly points towards an enrichment of favorable species. **Figure 3.13B** shows an enrichment of when the selection reached its completion at round 18. Different clones are starting to appear on multiple occasions.

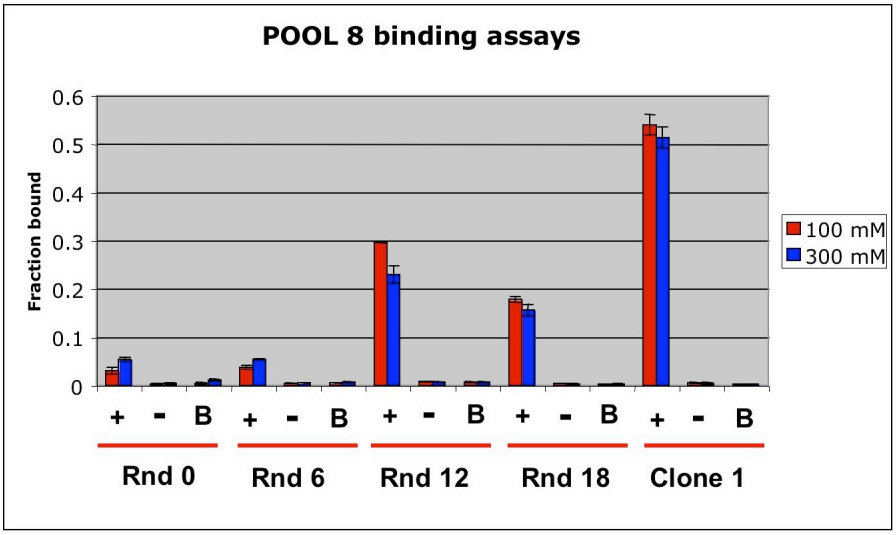


Figure 3.13A Affinity Trends of Pool 8 for Lysozyme. RNA from different rounds were assayed and incubated with protein Lysozyme (+), incubated with only buffer (-), or with pure Biotin (B). Two different wash buffers were analyzed. Binding assays washed with a buffer identical to the selection buffer conditions is shown in red bars, and a high salt buffer wash containing 3X more NaCl than the selection buffer is shown in blue bars.

Round 6 sequences

```
>G11 GGGAGCATTGCCCACTCATTTTCATTC-CGGCCGACGAGTAAGATATAAGCAAAGCCCACACAGAC---GTCACTAGTTCGCGTTGCTG
>E11 GGGAGCATTGCCCACTCATTTTCATTC-GAAAGGCTCTGGCTAGGAAGACTACAGTACACGGGTTTGG-GTCACTAGTTCGCGTTGCTG
>F9  GGGAGCATTGCCCACTCATTTTCATTC-GGCCCTAATGCGTACGGCGCTCGGTACACCTGAACACCCC-GTCACTAGTTCGCGTTGCTG
>F7  GGGAGCATTGCCCACTCATTTTCATTC-ACAAATGCTAATGGGGGGATGAAGATCGGGAACACGTTTA-GTCACTAGTTCGCGTTGCTG
>B12 GGGAGCATTGCCCACTCATTTTCATTC-GCACCCTCTACACCCTTATTGTCAGCCCAGTCCTAGCCCC--GTCACTAGTTCGCGTTGCTG
>D9  GGGAGCATTGCCCACTCATTTTCATTC-CCACACCCCCATAAAACGTATAACAAGCATGGTCCGGCC--GTCACTAGTTCGCGTTGCTG
>D7  GGGAGCATTGCCCACTCATTTTCATTC-CCGCATCAGCCCAACACAGATTGATTCAGGCCCGTCGGC--GTCACTAGTTCGCGTTGCTG
>A8  GGGAGCATTGCCCACTCATTTTCATTC-TCGGACTTGAGGTTGTGCGTGGCAGTTTATAGACGGGTCG-GTCACTAGTTCGCGTTGCTG
>H12 GGGAGCATTGCCCACTCATTTTCATTC-GGCGCGCAGGAGAAGTTACCAAGAACGGCCCGCTATTGTC-GTCACTAGTTCGCGTTGCTG
>A11 GGGAGCATTGCCCACTCATTTTCATTC-ACCGGGTTCAATAGGCAACTGGAGAAGGTATCGAAGTATG-GTCACTAGTTCGCGTTGCTG
>D11 GGGAGCATTGCCCACTCATTTTCATTC-TGACAAGGACCAAGTAGAGGCTAGGGTACATCTGGAGGTG-GTCACTAGTTCGCGTTGCTG
>E7  GGGAGCATTGCCCACTCATTTTCATTC-AGGAGCCCTCGCCACGTAACCACGTACATGCCATCTTGAC-GTCACTAGTTCGCGTTGCTG
>C12 GGGAGCATTGCCCACTCATTTTCATTC-GGCAAACCGGACCACTTTACTAGTAGCAACACCCCCGCCC-GTCACTAGTTCGCGTTGCTG
>B10 GGGAGCATTGCCCACTCATTTTCATTC-ATGTGCGGTGGAATGCATCCATCTAAGCTTCGGACGGTGG-GTCACTAGTTCGCGTTGCTG
>B11 GGGAGCATTGCCCACTCATTTTCATTC-GACCAGCCCAAATTGTTAGACCTTAAATCCTTGCCACCGCC GTCACTAGTTCGCGTTGCTG
>A12 GGGAGCATTGCCCACTCATTTTCATTC-CGGAGCCCGGCCCTCAACCAAGTATCAACTCCCACCCCGC-GTCACTAGTTCGCGTTGCTG
>H9  GGGAGCATTGCCCACTCATTTTCATTC-CACCCCGCCCCCAAGATATTTGAGAAGTCGAACCCCACAC-GTCACTAGTTCGCGTTGCTG
>H8  GGGAGCATTGCCCACTCATTTTCATTC-ACACCTGCGAGGGATCCCCAGGGGCCACAGGTAGCCGTAC-GTCACTAGTTCGCGTTGCTG
>H10 GGGAGCATTGCCCACTCATTTTCATTC-CAGGGTAGGCAGAAGAGGAAAGCGATCTTCACCGGTTGGC-GTCACTAGTTCGCGTTGCTG
>G9  GGGAGCATTGCCCACTCATTTTCATTC-AAACGAAACAGGCACCACGGGAATGGAGCACGGTATCTGG-GTCACTAGTTCGCGTTGCTG
>A9  GGGAGCATTGCCCACTCATTTTCATTC-TGTTTGCCGAATGGACACCCACCGCGACCTGCCCCCTGCGT GTCACTAGTTCGCGTTGCTG
>C11 GGGAGCATTGCCCACTCATTTTCATTC-GAACAGATCCCACCCAGCCGTGCTCACCTGGCCATCCCCC-GTCACTAGTTCGCGTTGCTG
>B8  GGGAGCATTGCCCACTCATTTTCATTC-AACAGTAGAGAAGACCGATGGATTGAACCCGTGATGTGGC-GTCACTAGTTCGCGTTGCTG
```

Round 12 sequences

```
>B7  GGAGCATTGCCCACTCATTTTCATTC-AGGACACGACTGCCACAGCCTCGGTCTAGTGCCCCCCGTT----GTCACTAGTTCGCGTTGCTG
>C7  GGAGCATTGCCCACTCATTTTCATTC-CGCGCCCCCGGATGCTACCGGATTACGACTACCAACTCCCCGGCGTCACTAGTTCGCGTTGCTG
>D7  GGAGCATTGCCCACTCATTTTCATTC-TGTGAGAACAAGGTCGATACGATGAAAACTGTTGGCGGGGTC--GTCACTAGTTCGCGTTGCTG
>E7  GGAGCATTGCCCACTCATTTTCATTC-CACGCATTTGTCAGATCACCCGTTGACACTGTCCCCGCCG----GTCACTAGTTCGCGTTGCTG
>B8  GGAGCATTGCCCACTCATTTTCATTC-TGGCTGATGCCCGTGCCCCTATGATGAAACGCCGCCCCTC----GTCACTAGTTCGCGTTGCTG
>C8  GGAGCATTGCCCACTCATTTTCATTC-GAGGTGGTACGAACACGGACTGCATCGTATGACACGCCCTG---GTCACTAGTTCGCGTTGCTG
>D8  GGAGCATTGCCCACTCATTTTCATTC-TCGGATGCCCTTCCCGCGCTGACTGTCCCCGTACGCCCTC----GTCACTAGTTCGCGTTGCTG
>F8  GGAGCATTGCCCACTCATTTTCATTC-CGCCCATCAGAGAGTACCCATCCCAGTAAATCCCCCGAGC----GTCACTAGTTCGCGTTGCTG
>B9  GGAGCATTGCCCACTCATTTTCATTC-ATGTCAATGCCGACTGAAATGGTTCCACCTTCCGCCCCCG----GTCACTAGTTCGCGTTGCTG
>C9  GGAGCATTGCCCACTCATTTTCATTC-AGGATGAGAGCGTACATGGTAGAGACAACGAGCCGGCGTG----GTCACTAGTTCGCGTTGCTG
>D9  GGAGCATTGCCCACTCATTTTCATTC-GGGCTGGACCGTTGACTAGATACGATATTAGGAGCCGGGG----GTCACTAGTTCGCGTTGCTG
>E9  GGAGCATTGCCCACTCATTTTCATTC-TTCGGGGTAACACAAGATAACAAGGGAGCGATCGGGATCA----GTCACTAGTTCGCGTTGCTG
>F9  GGAGCATTGCCCACTCATTTTCATTC-TAGGGTTAACATTCAAAGGAACAGGTAATGGCGGGTCCAG----GTCACTAGTTCGCGTTGCTG
>G9  GGAGCATTGCCCACTCATTTTCATTC-AAGAACGATGGGAGAAGGAAATTAATGAACTGGCCGCAAA----GTCACTAGTTCGCGTTGCTG
>B10 GGAGCATTGCCCACTCATTTTCATTC-CACGCACCCAGATGCTAACTGAAACGCACCACTCGCCGGC----GTCACTAGTTCGCGTTGCTG
>B12 GGAGCATTGCCCACTCATTTTCATTC-TGCCGAACCACAAGTGACCCCCGCACTTGTAGACCGACGCGGGTGTCACTAGTTCGCGTTGCTG
>C12 GGAGCATTGCCCACTCATTTTCATTC-AACTTCATGCCCAAACGCTACGCATTCTACCTCCCCCCGC----GTCACTAGTTCGCGTTGCTG
>E12 GGAGCATTGCCCACTCATTTTCATTC-AGCGCAATCAAATAATGTGAACGACCCCCTCGGCCCATAC----GTCACTAGTTCGCGTTGCTG
```

Round 18 sequences

```
>C2  GGGAGCATTGCCCACTCATTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>B2  GGGAGCATTGCCCACTTTTTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>G1  GGGAGCATTGCCCACTCATTT-CATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>E2  GGGAGCATTGCCCACTCTTTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>F4  AGGAGCATTGCCCACTCTTTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAACGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>H6  GGGAGCATTGCCCACTCATTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>A6  GGGAGCATTGCC-ACTCATTTTCATTC-GTCGATAAGGACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>H3  GGGAGCATTGCCCACTCATTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>G6  GGGAGCATTGCCCATTTATTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>F3  GGGAGCATTGCCCACTCATTTTCATTC-GTCGATA-GTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>C1  GGGAGCATTGCCCACTCATTTTCATTC-GTGGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG

>E1  GGGAGCATTGCCCACTCATTTTCATTC-CCAAGGCATTGATGATGGCGATCTCAAGAATATGACGGCGGTG-GTCACTAGTTCGCGTTGCTG
>G5  GGGAGCATTGCCCACTCATTTTCATTC-CCAAGGCATTGATGATGGCGATTAAGAAATGACGGCGGTG----GTCACTAGTTCGCGTTGCTG
>G2  GGGAGCATTGCCCACTCTTTTTCATTC-CCAAGGCATTGATGATGGCGATTAAGGTATGACGGCGGTG----GTCACTAGTTCGCGTTGCTG
>H4  GGGAGCATTGCCCACCTCATTTCATTC-CCAAGGCATTGATGATGGCGATTAAGGTATGACGGCGGTG----GTCACTAGTTCGCGTTGCTG
>H2  GGGAGCATTGCCCACTCATTTTCATTC-CCAAGGCATTGATGATGGCGATTAAGGTATGACGGCGGTG----GTCACTAGTTCGCGTTGCTG

>B6  GGGAGCATTGCCCATTTATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTTGACACCGGCCCCC---GTCACTAGTTCGCGTTGCTG
>B4  GGGAGCATTGCCCATTCATGTTCATTC-AGCCAAGCCGCCCCCTCCGAACTCACTCGACACCGGCCGCC----GTCACTAGTTCGCGT-GCTG
>E3  GGGAGCATTGCCCTTTCATTTTCATTC-AGCCAAGCCGCCCCCTCCGAACTCACTCGACACCGGCCCCC---GTCACTAGTTCGCGTTGCTG
>D1  -GGAGCATTGCCCACTCATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTCGACACCGGCCCCC----GTCACTAGTTCGCGTTGCTG
>A3  GGGAGCATTGCCCATTTATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTCGACACCGGCCCCC----GTCACTAGTTCGCGTTGCTG
>F1  GGGAGCATTGCCCATTCATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTCGACACCGGCCCCC----GTCACTAGTTCGCGTTGCTG
>C5  GGGAGCATTGCCCACTCATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTACGACACCGGCCCCC---GTCACTAGTTCGCGTTGCTG

>E5  GGGAGCATTGCCCACTCATTTTCATTC-AGCGTAAATGGGCACTTACGAAATAACAAGGTCGGTGAGG----GTCACTAGTTCGCGTTGCTG
>G3  GGGAGCATTGCCCACTCATTTTCATTC-AGCGTAAATGGGCACTTACGAAATAACACGGTCGGTGAGG----GTCACTAGTTCGCGTTGCTG
>D3  GGGAGCATTGCCCACTCATTTTCATTC-AGCGTAAATGGGCACTTACGAAATAACACGGTCGGTGTGG----GTCACTAGTTCGCGTTGCTG

>B1  GGGAGCATTGCCCACTCATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCTG
>B3  GGGAGCATTGCCCATTCATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCTG
>G4  GGGAGCATTGCCCATTCATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCTG
>C3  GGGAGCATTGCCCATTCATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCAG
>H1  GGGAGCATTGCCCACTCATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCTG

>C4  GGGAGCATTGCCCATTTATTTTCATTC-AGTACCTGACCAGCACCCGTAGGGATCCTGACCCGCCGCC----GTCACTAGTTCGCGTTGCTG
>E4  GGGAGCATTGCCCTTTCATTTTCATTC-AGTACCTGACCAGCACCCGTAGGGATCCTGACCCGCCGCC----GTCACTAGTTCGCGTTGCTG

>D2  GGGAGCATTGCCCACTCATTTTCATTC-CAGAACCCCCCCTCGAGAAAGCTACACCTTAGTCCCCCGGC----GTCACTAGTTCGCGTTGCTG
>D4  GGGAGCATTGCCCACTCATTTTCATTC-CAGACCCCCCCCTCGAGAAAGCTACACCTTAGTCCCCCGGC----GTCACTAGTTCGCGTTGCTG

>F2  GGGAGCATTGCCCATCTATTTTCATTC-CACGCACCCAGATGCTAACTGAAACGCACCACCCGCCGGC----GTCACTAGTTCGCGTTGCTG
>A5  GGGAGCATTGCCCACTCATTTTCATTC-CACGCACCCAGATGCTAACTGAAACGCACCACCCGCCGGC----GTCACTAGTTCGCGTTGCTG
>F5  GGGAGCATTGCCCATTCATTTTCACTC-GCATCTGAACAAGGATACGTAGCGTTTAGAACCCCCCGCC----GTCACTAGTTCGCGTT----
>B5  GGGAGCATTGCCCACTCATTTTCATTC-GCACACACAGCAATGATAGTTCATCAAGACGAAGCCGGGG----GTCACTAGTTCGCGTTGCTG
>A4  GGGAGCATTGCCCATTATTTTCATTC-AGGATGAGAGCGTACATGGTAGAGACAACGAGCCGGCGTG----GTCACTAGTTCGCGTTGCTG
>E6  GGGAGCATTGCCCACTCATTTTCATTC-TAAGAGGGGGGTCCATCGCAAACAGAACACGGATGTCGGCA---GTCACTAGTTCGCGTTGCTG
```

Figure 3.13B. Pool 6 Sequence Clones from Rounds 6, 12, and 18. N40 pool sequences isolated from Pool 1 selection. The primer binding regions are colored maroon. Random regions are in black. Random regions that are highlighted signify identical sequences. Underlined sequences show common motifs among the aptamer clones.

ASSESSING THE ABSENCE OF CONTAMINATION IN THE SELECTIONS

Phylogenetic analysis of the round 18 sequences show that all 8 different pool aliquots have been selected successfully without cross contamination across the different pool aliquots during the selection process as shown in **Figure 3.14**. Identical sequences within the same pool aliquot were input only once into the analysis. The phylogenetic analysis shows that no two sequences from the different pools shared the same branch as it would be expected from identical sequences. The number of sequences was reduced in order to provide an intelligible graphical representation. Reducing cross contamination is one of the greatest challenges for automated selection. The lack of enclosure of the robotic platform makes contamination a likely event during the run. Thorough cleaning after each run and the use of filter systems that removes nucleic acids aerosols reduced this phenomenon. The mechanism of robotic selections had to be modified in order to overcome the contamination issues that were observed with the N30 pool for the N40 the selections.

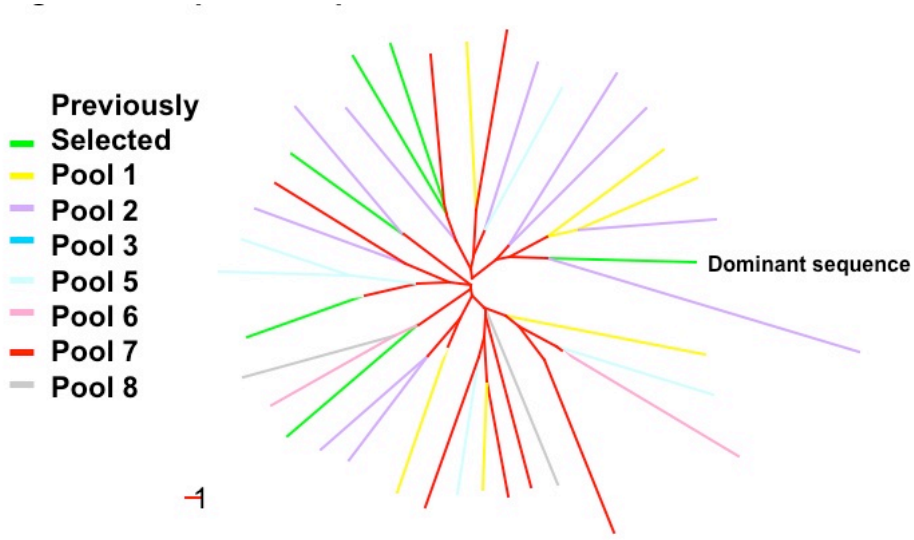


Figure 3.14. Phylogenetic Analysis of Round 18 Sequences. The phylogeny is constructed by comparing the hamming distance of the random region among the isolated aptamers and plotted using the DNAdist software from the PHYLIP software package [18]. The lack of overlap among the phylogenetically analyzed sequences indicates that there is no cross contamination among the 8 pools selected in parallel. The number of sequences used has been reduced in this graph. Previously isolated aptamers are incorporated into the analysis to show that no previously isolated aptamers have been isolated in the new selections.

Conclusions

A automated *in vitro* selection system capable of performing selections in by proceeding through six rounds of selection before fresh reagents are required (usually a 18 – 20 hour period) has been designed in a modular format, utilizing as many off-the-shelf components as possible so that we can quickly replace broken subsystems, or rapidly build new workstations. Despite the modularity of this system, it remains highly configurable, allowing a wide variety of selection conditions and procedures. As versatile as the system seems, the issue of cross contamination within the system can make it less than desirable to be used. With a few modifications, 8 pool parallel selections had been

carried out successfully. Not only were the problems resolved further validating the utility of the robotic platforms; it also opened many different areas of research that can be accessed with this technology.

These selections were not only successful at addressing the contamination issue; it also gave us a glimpse at the degree of variation in the behavior of different pools when they contain different unique sequences. Based on the binding data, it can be inferred the extent of pool enrichment that has occurred during a round of selection. A question that arise from this experiment is that if there is such diversity when there is such difference in complexity is how efficient are *in vitro* selections at isolating the best binders? Many aspects such as effects of priming regions [19], type of selection buffer [20-25], length of pool, and target immobilization are known to how they can affect the outcome of selection. But the effect of different complexity of the pool has never been tapped into experimentally. Chapter 4 is devoted to analyzing the outcomes of the individual aptamers isolated from each pool and compared to one another.

The aspect of theoretical biology and biological systems modeling on computer workstations might have seemed only limited to the field of bioinformatics. However, with the ability to generate such a vast amount of data in a short amount of time experimentally, studying evolutionary patterns is not far from reach. Chapter 4 is provides a glimpse at the extent that automated selection can be used to probe a small area of theoretical biology that was reserved to the selected few.

THE FUTURE OF ROBOTICS

Different students in the lab throughout the years have investigated adaptation of

this technology to a larger robotic worksurface in the past. Having the cross contamination issues addressed, the robot can be further outfitted to select very closely related targets. One of the main interests in the lab is to automate cell-surface selections. Most tumor mammalian cells possess highly similar surface properties; nixing cross contamination is critical for successful selections. I have currently worked out the particulars such as components that are needed for automated cell surface selections and a High School student Daniel Winkler is in the process of incorporating them into the Biomek 2000 and Tecan Genesis platforms.

MATERIALS AND METHODS

NEW SELECTION POOL DESIGN

In traditional primer and pool design, it is commonplace to have 'strong' (S) bases, such as ---GGC towards the primer ends to increase priming efficiency [26]. An important trait of automation pools that makes them different from normal selection pools is that they have 'weak' (W) 3' primer binding ends. An important feature of robotic selections is by-passing of PAGE purifications, and the uniform number of PCR cycles for numerous of rounds of selection, it is crucial to design a pool that is able to undergo an infinite number of amplification steps without the generation of parasites. Since amplification efficiency is one of the crucial aspects of automated selections, having a SWW 3' primer ends that hybridize poorly will help this aspect without the generation of unwanted amplicons that can occur in ends that are characterized by WSS clamps [19].

POOL SYNTHESIS AND PURIFICATION

One μmole syntheses of pool template strand and primers were synthesized on an Expedite DNA synthesizer (Applied Biosystems, Foster City, CA). N40 pool (GGG AGC ATT GCC CAC TCA TTT TCA NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN N TCA CTA GTT CGC GTT GCT GAA); 42.40 (GAT AAT ACG ACT CAC TAT AGG GAG CAT TGC CCA CTC ATT TTC); 20.40 (TTC AGC AAC GCG AAC TAG TGA); where "N" represents the random region of the pool, where all four phosphoramidites are injected simultaneously. Because each base has a different coupling efficiency, the "N" mixture is prepared with A:C:G:T phosphoramidites ratios of 3:3:2:2.4, respectively [27]. The synthesized nucleic acid was deprotected by incubation of the synthesis resin in 1 ml of ammonia hydroxide at 50ºC for 16 hours. A 10x volume of n-butanol was added and oligos were allowed to precipitate at -20ºC overnight. The oligos were spun down, ethanol rinsed, and then gel purified on polyacrylamide gels. One-third of the pool syntheses were purified per each 10 x 12 x 0.15 cm 8% acrylamide gel. One-half of the primer syntheses were purified per each 10 x 12 x 0.15 cm 12% acrylamide gel. After running the gels for approximately two hours at 250 V (until the bromophenol blue loading dye reaches the bottom of the gel), the nucleic acid was visualized by placing the gel on a thin-layer chromatography plate and UV shadowing [28].Visual bands containing the appropriate mass of nucleic acid were excised with a clean razor blade. Gel fragments were eluted in pure water, approximately 30 ml per 10 x 12 x 0.15 cm gel excision. After overnight elution at 37ºC on a rotator, the water nucleic acid mixture was ethanol precipitated with sodium acetate, and quantitated by UV spectroscopy. The purified N30 synthesis was determined to

contain a total of 2.44x1016 strands.

N40 LIBRARY COMPLEXITY DETERMINATION

During nucleic acid library synthesis, a substantial portion of the pool created by the synthesizer is not extendable my polymerases. Backbone lesions, incomplete de-protection of the bases, and deletions are the artifacts that arise during synthesis affecting pool extension [27]. Pool extension assays are crucial for determining the complexity of the pool and for determining the fraction of the pool that is amplifiable with a polymerase. The assay involves the radiolabeling of 40 pmol of the pool primer (20.40) in 10 uL of kinase reaction with 5X forward polynucleotide kinase reaction buffer, γ-^{32}P ATP (>7,000 Ci/mmol) (ICN Biomedicals, Irvine, CA), and T_4 polynucleotide kinase (Invitrogen, Carlsbad, CA). The reaction is incubated at 37ºC for 1 hour and phenol:cholorform extracted. The primers are separated from the organic phase using a PhaseLock tube (Eppendorf, Hamburg, Germany). The final step involves the removal of free nucleotides by size-exclusion chromatography with Centri-Spin 20 microcentrifuge columns (Princeton Separations, Adelphia, NJ).

Both a DNA polymerase and a reverse transcriptase (RT) are used to test for pool extension. RTs are more efficient at extending the pools because of its error-prone nature, meaning that they are able to overcome chemically modified or damaged base strands [29-31]. *Taq* DNA polymerase (Promega) and SuperScript III reverse transcriptase (Invitrogen) are the two enzymes to evaluate pool extension. For the RT reaction, it involves incubating an excess amount of the pool (50 pmol) with a limiting amount of the labeled primer (10 pmol) with 10 nmol of dNTPs and water in a 10 uL

reaction, heat denatured at 70ºC for 5 minutes and chilled. First strand buffer (4 uL), 0.1M DTT (2 uL), 1 unit of RT enzyme, and water are added to bring the reaction to a final volume of 20 uL. The reaction is incubated at 42ºC for 1 hour. For the *Taq* DNA polymerase reaction, the radiolabeled primer is annealed to the ssDNA pool by adding 10 pmol of the primer with 50 pmol of the purified pool. Additional 10 nmol of dNTPs and water added to bring the volume to 10 uL followed by a denaturing step at 70ºC for 5 minutes and chilled. PCR buffer added to the chilled reaction along with 1 unit of the *Taq* enzyme. The reaction is incubated at 72ºC for 1 hour. Once both reactions are completed, 5 uL aliquots of each reaction is added to an 5 uL of 2X stop dye and denatured at 70ºC for 3 minutes. The sample are then analyzed on an 8% PAGE gel and imaged (**Figure 3.15)**.

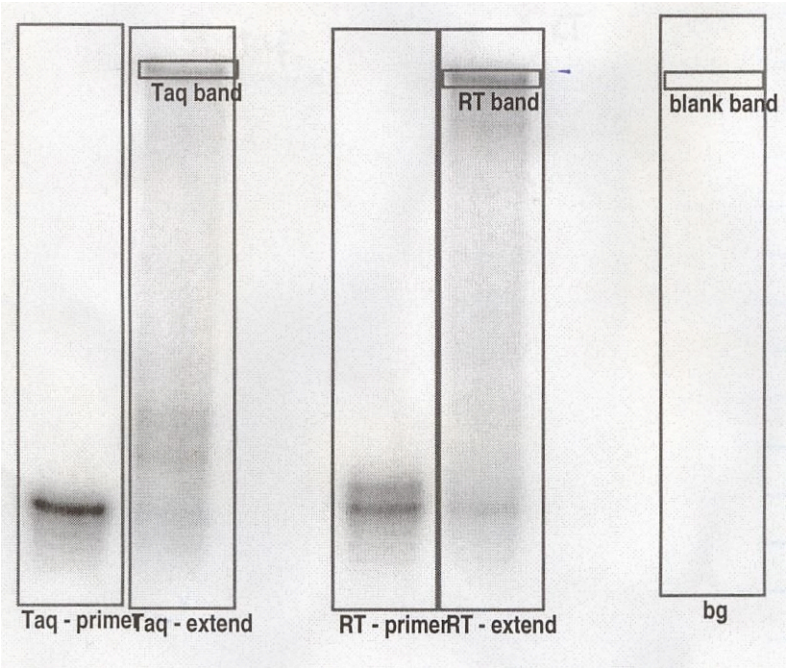


Figure 3.15. Extension Analysis of Synthesized Pool. The ssDNA pool is analyzed for extension using radiolabeled primers. A DNA polymerase and Reverse Transcriptase and used and compared to each other. One column is the primer itself, the second column is

the primer incubated with the pool.

POOL AMPLIFICATION FOR SELECTIONS

Before each new pool can undergo selections, a small scale PCR is crucial to optimize the reaction conditions. These are needed to ensure that the new pools are not over-amplified, which is a critical aspect of successful selections [27]. Small-scale reactions are almost a mirror image of the large-scale reactions carried out in PCR thin well microplates (Bio-Rad Laboratories, Hercules, CA). The initial amount of template required to start a reaction follows the assumption that a traditional 100 uL PCR reaction will yield at most 1 ug of products. The initial goal is to obtain 10 copies of each viable nucleic acid sequence and my starting condition for each pool is $1x10^{14}$ unique viable fully extendable sequences from the synthesis. The mass of 1 double-stranded copy of 10^{14} molecules is $1.2x10^{-5}$ g; hence for 10 double stranded copies of the N40 DNA will be $1.2x10^{-4}$ g. If each PCR reaction would yield 1 ug of product, to obtain the mass for 10 copies would require 12 mL of PCR reaction (120 reactions of 100 uL). Since for each pool that is going to be tested has $1x10^{14}$ unique sequence in 100 uL volume, each small-scale PCR reaction will contain 0.83 uL of the ssDNA template, 4 pmol of each primer (20.40 and 42.40) and 20 nmol of dNTPs. Once the PCR conditions are optimized, the large-scale amplification of the ssDNA pools can take place to generate the library necessary to initiate the selections.

For the purpose of this study, extreme caution had to be conducted to avoid cross-contamination across the 8 different pools. Once the newly synthesized pool has been divided into 8 aliquots single aliquots, in no time during the small-scale or large-scale PCR amplification have any two aliquots been worked on at the same time. During this sensitive time, each pool was worked individually under a hood that is that is cleaned

with bleach, RNAse Zap solution (Applied Biosystems, Foster City, CA), and ethanol before and after each manipulation of the pool. This is to ensure the integrity of the pool.

Once the each aliquot is PCR amplified, all the reactions are pooled and ethanol precipitated with 0.3 M of sodium acetate. The DNA pellets are then suspended in water and checked for integrity on a 4% agarose gel. Once the dsDNA copies are obtained, they will be transcribed into the RNA form. Large-scale transcriptions are conducted for each pool separately. The 1 mL transcription reactions include 7.5 mM of each nucleotide (ATP, CTP, GTP and UTP), 10 mM of DTT, 1X T7 transcription buffer obtained from AmpliScribe kits (Epicentre, Madison, WI) and the dsDNA previously amplified. The reactions are incubated at 42ºC for 4 hours. All traces of DNA in the transcription is removed by the addition of DNase I and incubated at 37ºC for an additional hour. The RNA is purified on an 8% PAGE gel, eluted and ethanol precipitated.

TARGET PREPARATION

Lysozyme purified from hen egg white was purchased from Sigma-Aldrich (St. Louis, MO). The enzyme was suspended at a 1 mg/mL concentration in the selection buffer (see below), and chemically biotinylated using sulfo-NHS-LC-biotin (Pierce, Rockford, IL). The biotinylation reaction occurred at a 10× molar ratio of biotinylating reagent to enzyme, and proceeded for 2 h on ice. Unincorporated biotin was removed via a 10DG chromatography desalting column (Bio-Rad, Hercules, CA). The biotinylated lysozyme was then captured by magnetic strepavidin-coated Dynabeads (Dynal, Lake Success, NY).

SELECTION WORKSURFACE

Figure 3.16A [32]. Image of the Selection System Used for Selections. The system is outfitted with a thermalcycler with an automated lid integrated and controlled through the computer system that runs the robot. An enzyme cooler allows the enzymes to remain active for over 18 hours. The carousel functions to feed the system with new pipet tips.

The selections are carried out on a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, Fullerton, CA) a shown in **Figure 3.16A** [32]. The robot

is integrated with a PTC-200 thermal cycler (MJ Research, Waltham, MA), a multiscreen vacuum filtration manifold (Millipore, Bedford, MA), and an enzyme cooler engineered in the laboratory by Tim Riedel. A Stacker Carousel (Beckman Coulter) feeds fresh pipette tips to the workstation as needed. **Figure 3.16B** [32] is a diagram of the worksurface and the placement of the plates that play a role in the selection process. A unique aspect of the N40 selection is the integration of a plate shaker in the location of the target plate. Since the magnetic beads tend to cement to the bottom of the plate, a shaking motion keeps the beads mobile thus reducing the necessity of constantly pipette just to keep the mixture well mixed. Pipetting constantly may cause the attachments of the beads and targets to the tips, which can lead to potential loss of RNA binding species.

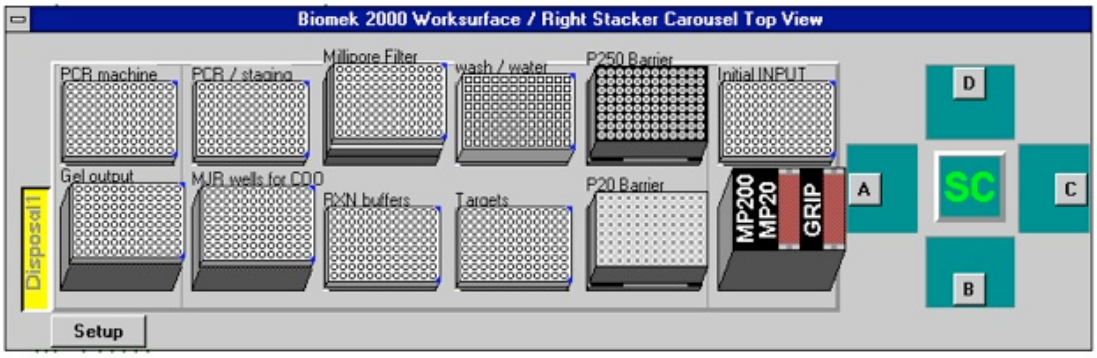


Figure 3.16B [32]. Schematic of Biomek 2000 worksurface. The relative positions of the microwell plates and labware that are involved in the aptamer selection process are shown.

SELECTION PROCESS

Colin Cox and Letha Sooterhad had previously designed the mechanical manipulations involved in Biomek 2000 protein selections [10]. In summary, the

principal differences between manual and automated selections are the protein target preparation. In automated selections targets are initially biotinylated and subsequently conjugated to magnetic beads after initial incubation of the nucleic acid pool and the protein target. The basic steps is illustrated in **Figure 3.17** [8]. The modification involved in the N40 selections is the first incubation step. Instead of introducing the nucleic acid pool to biotinylated targets conjugated to beads, they are introduced to biotinylated targets prior to their conjugation to streptavidin coated magnetic beads. The rationale for this step is to reduce the non-specific binding of nucleic acids to the beads. The eight RNA pools are carried through 18 rounds of *in vitro* selection simultaneously against the hen egg white Lysozyme using the Beckman Biomek 2000 workstation. In the first round of selection, the amount of RNA of each pool applied is 5 ug (ca. $1.1x10^{14}$ sequences). Thereafter, approximately one-fifth of the preceding RNA transcription reaction is applied to the solution of biotin-lysozyme to begin the next round of selection.

The very first step of the selection involves incubating the RNA with the biotinylated lysozyme in selection buffer (20mM Tris (pH 7.5), 100mM NaCl, 5mM $MgCl_2$). Streptavidin-derivatized Dynabeads (Dynal Biotech, Brown Deer, WI) are used to 'rescue' or conjugate the biotin-lysozyme and any bound RNA. The bead, target, and RNA mixture is then filtered through a Millipore HV (PVDF) filter on a vacuum manifold to partition the lysozyme binding RNA from the unbound RNA. The winning RNA species are eluted from the target through a temperature denaturing step and then amplified through a reverse transcription coupled PCR (RT–PCR) reaction followed by a transcription reaction. The RT–PCR reaction buffer contains 10 mM Tris (pH 8.4), 50 mM KCl, 1.5 mM MgCl2, 0.2 mM dNTPs, 5% acetamide, 0.05% Nonidet P40, and 0.5

mM each of the primers). RT–PCR enzyme mixture contains 5 U of SuperScript III enzymes (Invitrogen), and 0.2 U of Display Taq (Display Systems, Vista, CA), 50% glycerol, 10 mM Tris (pH 8.4), 50 mM KCl, and 1.5 mM $MgCl_2$. A fraction of the RT–PCR mixture serves as template for the following transcription reaction to generate the enriched RNA species for the next round of selection. The transcription buffer has 40 mM Tris (pH 7.9), 26 mM MgCl2, 5 mM dithiothreitol (DTT), and 2.5 mM of each NTP. The transcription enzyme mix contained 40 units of RNasin (Promega, Madison, WI) ribonuclease inhibitor and 100 U of T7 RNA polymerase (Stratagene, La Jolla, CA). The selection progress is evaluated at every round by adding α-^{32}P radiolabeled ribo- and deoxyribonucleotides to the amplification reactions buffers and archiving small aliquots (10 uL) of the reactions to be resolved in 8% acrylamide denaturing gel. The gels are dried visualized using a PhosphorImager SI (Amersham Pharmacia Biotech).

RNA POOL ASSAYS

Individual pools were assayed for binding activity by incubating equimolar amounts of the isolated RNAs to Lysozyme in selection buffer. Nitrocellulose filter-binding assays were employed to determine the fraction of aptamer bound to Lysozyme [33] using A minifold I filtration manifold (Schleicher and Schuell, Keene, NH). A sandwiched nitrocellulose (Protran from Schleider & Schuell) and nylon membrane (Hybond N+ from Amersham Pharmacia Biotech) is assembled in the modified minifold attached to a vacuum. The binding reactions are filtered through the assembled apparatus and washed with 300 uL of wash buffer. The amount of radiolabeled RNA captured on the nitrocellulose is quantitated using a PhosphorImager SI (Amersham Pharmacia

Biotech). The fraction of RNA captured is calculated by the following equation:

$$\text{Fraction of RNA bound} = \frac{\text{RNA captured on nitrocellulose}}{\text{RNA on nitrocellulose + RNA on nylon}}$$

The assays for the different pools were carried out on the same day to reduce the amount of variation during sample manipulations. Clone 1 used in the assay was the previously selected lysozyme aptamer [8] serving as the positive control. RNA was also incubated with the Dynalbeads and assayed to ensure that the aptamers selected were not bead binders.

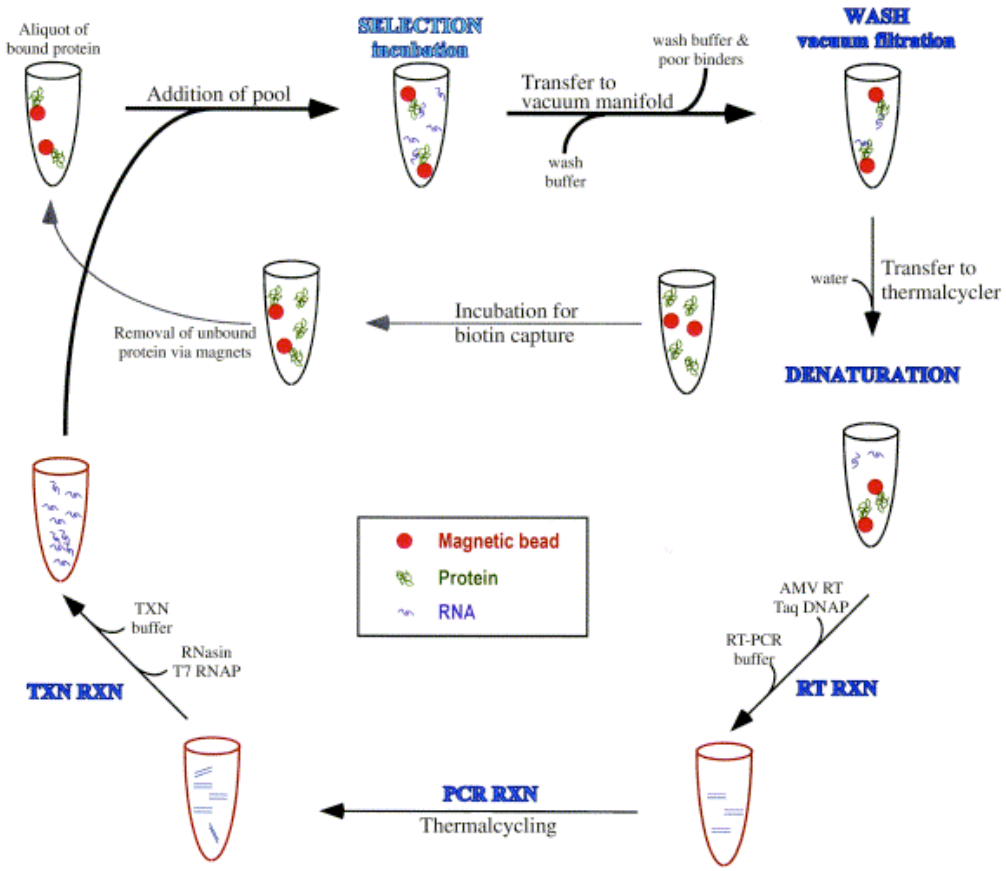


Figure 3.17. Robotic Aptamers Selections Against Protein Targets [8]. This diagram depicts the steps that are involved in one single round of selection that takes place on the

Biomek 2000. The red dots represent the magnetic beads coated with streptavidin, which aids in the conjugation and capture of the biotinylated protein targets with nucleic acid binders. The initial steps involves the addition of the RNA pool to a protein target captured on magnetic beads followed by filtration washing of the beads, elution of selected binding species, and amplification to regenerate RNA for the start of the next round.

REFERENCES

1. **Boland, E.J., et al., *Automation of the Maxam-Gilbert chemical sequencing reactions.* Biotechniques, 1994. 16(6): p. 1088-92, 1094-5.**

2. **Civitello, A., Richards, S., Gibbs, RA., *A simple protocol for the automation of DNA cycle sequencing reactions and polymerase chain reactions.* DNA Seq., 1992. 3(1): p. 17-23.**

3. **King, R.D., et al., *Functional genomic hypothesis generation and experimentation by a robot scientist.* Nature, 2004. 427(6971): p. 247-52.**

4. **Benkestock, K., et al., *Automated nano-electrospray mass spectrometry for protein-ligand screening by noncovalent interaction applied to human H-FABP and A-FABP.* J Biomol Screen, 2003. 8(3): p. 247-56.**

5. **Chapman, T., *Lab automation and robotics: Automation on the move.* Nature, 2003. 421(6923): p. 661, 663, 665-6.**

6. **McPherson, J.D., et al., *A physical map of the human genome.* Nature, 2001. 409(6822): p. 934-41.**

7. **Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. 291(5507): p. 1304-51.**

8. **Cox, J.C. and A.D. Ellington, *Automated selection of anti-protein aptamers.* Bioorg Med Chem, 2001. 9(10): p. 2525-31.**

9. **Cox, J.C., et al., *Automated acquisition of aptamer sequences.* Comb Chem High Throughput Screen, 2002. 5(4): p. 289-99.**

10. **Cox, J.C., P. Rudolph, and A.D. Ellington, *Automated RNA selection.* Biotechnol Prog, 1998. 14(6): p. 845-50.**

11. **Sooter, L.J., et al., *Toward automated nucleic acid enzyme selection.* Biol Chem, 2001. 382(9): p. 1327-34.**

12. **Grundy, F.J. and T.M. Henkin, *From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements.* Crit Rev Biochem Mol Biol, 2006. 41(6): p. 329-38.**

13. **Malmgren, C., et al., *An antisense/target RNA duplex or a strong intramolecular RNA structure 5' of a translation initiation signal blocks ribosome binding: the case of plasmid R1.* Rna, 1996. 2(10): p. 1022-32.**

14. **Schmitt, T. and N. Lehman, *Non-unity molecular heritability demonstrated by continuous evolution in vitro.* Chem Biol, 1999. 6(12): p. 857-69.**

15. **Harada, K. and L.E. Orgel, *In vitro selection of optimal DNA substrates for T4 RNA ligase.* Proc Natl Acad Sci U S A, 1993. 90(4): p. 1576-9.**

16. **Hesselberth, J., et al., *In vitro selection of nucleic acids for diagnostic applications.* J Biotechnol, 2000. 74(1): p. 15-25.**

17. **Hesselberth, J.R., et al., *In vitro selection of RNA molecules that inhibit the activity of ricin A-chain.* J Biol Chem, 2000. 275(7): p. 4937-42.**

18. **Felsenstein, J., *An alternating least squares approach to inferring phylogenies from pairwise distances.* Syst Biol, 1997. 46(1): p. 101-11.**

19. **Crameri, A. and W.P. Stemmer, *10(20)-fold aptamer library amplification without gel purification.* Nucleic Acids Res, 1993. 21(18): p. 4410.**

20. **Clark, S.L. and V.T. Remcho, *Electrochromatographic retention studies on a flavin-binding RNA aptamer sorbent.* Anal Chem, 2003. 75(21): p. 5692-6.**

21. **Cowan, J.A., et al., *Recognition of a cognate RNA aptamer by neomycin B: quantitative evaluation of hydrogen bonding and electrostatic interactions.* Nucleic Acids Res, 2000. 28(15): p. 2935-42.**

22. **Dapic, V., et al., *Biophysical and biological properties of quadruplex oligodeoxyribonucleotides.* Nucleic Acids Res, 2003. 31(8): p. 2097-107.**

23. **Draper, D.E., *A guide to ions and RNA structure.* Rna, 2004. 10(3): p. 335-43.**

24. **Parrott, A.M., et al., *RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour.* Nucleic Acids Res, 2000. 28(2): p. 489-97.**

25. **Zimmermann, G.R., et al., *Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer.* Rna, 2000. 6(5): p. 659-67.**

26. **Chen, B.Y., H.W. Janes, and S. Chen, *Computer programs for PCR primer design and analysis.* Methods Mol Biol, 2002. 192: p. 19-29.**

27. **Pollard J., S.D.B., and A.D. Ellington, *Design, synthesis, and amplification of DNA pools for in vitro selection*, in *Current Protocols in Nucleic Acid Chemistry*, e.a. S.L. Beaucage, Editor. 2000, John Wiley & Sons: New York. p. 9.2.1-9.2.23.**

28. **Ellington, A.D., *Purification of Oligonucleotides Using Denaturing Polyacrylamide Gel Electrophoresis*, in *Short protocols in molecular biology.*, e.a. F. Ausubel, Editor. 1997, John Wiley & Sons, Inc.: New York.**

29. **Bakhanashvili, M. and A. Hizi, *Fidelity of the RNA-dependent DNA synthesis exhibited by the reverse transcriptases of human immunodeficiency virus types 1 and 2 and of murine leukemia virus: mispair extension frequencies.* Biochemistry, 1992. 31(39): p. 9393-8.**

30. **Battula, N. and L.A. Loeb, *On the fidelity of DNA replication. Characterization of polynucleotides with errors in base-pairing synthesized by avian myeloblastosis virus deoxyribonucleic acid polymerase.* J Biol Chem, 1975. 250(12): p. 4405-9.**

31. **Mendelman, L.V., J. Petruska, and M.F. Goodman, *Base mispair extension kinetics. Comparison of DNA polymerase alpha and reverse transcriptase.* J Biol Chem, 1990. 265(4): p. 2338-46.**

32. **Lee, J.F., Cox, J.C., Collett, JR., Ellington, AD, *Exploring Sequence Space Through Automated Aptamer Selections.* Journal of the Association for Laboratory Automation, 2005. 10(4): p. 213-218.**

33. **Jhaveri, S.a.E.A., *In vitro selection of RNA aptamers to a protein target by filter immobilization*, in *Current Protocols in Nucleic Acid Chemistry*, e.a. S.L. Beaucage, Editor. 2000, John Wiley and Sons, Inc.: New York. p. 9.3.1-9.3.25.**

CHAPTER 4: PROBING APTAMER FITNESS LANDSCAPES WITH AUTOMATED *IN VITRO SELECTIONS*

INTRODUCTION

FITNESS LANDSCAPES

Fitness has very different meanings. In population genetics, fitness describes the ability of a genotype to survive from one generation to another. Some evolutionary biologists use fitness to describe phenotype adaptation to the environment [1]. Overall, fitness is subjective depending on the feature the end user is trying to test and describe. In our study, we define fitness as the ability of the aptamer to evolve its binding to the target lysozyme.

The concept of adaptive walks on fitness landscapes has captivated evolutionary biologists ever since Sewall Wright introduced it in 1932 [2]. According to Wright, to improve fitness, a fitness peak must be reached within a population by "climbing" a landscape. For theoretical biologists, understanding the topography of the landscapes has been of great interest. An optimal system is defined as the ability of the system to arrive at a global maximum in a multidimensional fitness space, and finding this optimum is also known as a combinatorial optimization [3]. Since biological systems are considered to be highly complex, it is assumed that their fitness landscapes are often extremely rugged, especially when their parameter involves their genotype or sequence space [4]. The phenotype of a system dictates its viability and fitness landscapes are a representation of this property on sequence space. Hence fitness is a multidimensional

function of sequence space.

The nature of the landscape that correlates molecular sequence to molecular function has been the focus of a number of evolutionary theorists. These theoretical studies have primarily been modeled *in silico* due to the vast amounts of data needed that were deemed unattainable through the traditional wet-lab approach. Theoretical biologists have for many years speculated on the nature of fitness landscapes, such as how sequence space maps to any of a number of different phenotypes (including folding, catalysis, and substrate specificity). For example, Schuster and his co-workers have used *in silico* models to exhaustively examine how nucleic acid sequences map to particular secondary structures [5, 6]. These authors have come to several remarkable conclusions, including the notion that most available secondary structures can be found in a relatively small subset of sequence space. It is therefore not necessary to explore the entire sequence space to attain the best (or evolutionarily successful) structure. Such findings have important implications and may be tested experimentally.

ROLES OF *IN VITRO* EVOLUTION AND SEQUENCE SPACE ANALYSIS

With the advent of selection and amplification techniques, conducting directed evolution provides detailed insight into the dynamics of evolving *in vitro* systems. For example, the information that can be obtained from evolving RNAs can provide specific biological behaviors associated with certain sequences, which in turn translates into the effect or association between phenotype (secondary structure) and genotype (primary structure). The *in vitro* evolution of ribozymes has provided better understanding on how natural RNA catalysts evolve. Previous *in vitro* isolation of a group of RNA-cleaving

ribozymes with a motif that is similar to the hammerhead ribozymes provides evidence that self-cleaving ribozymes have undergone convergent evolution in nature [7].

The Bartel lab used *in vitro* evolution of ribozymes to prove the notion that RNA folds can evolve without the requirement of inactive intermediates [8]. They started with a class III RNA ligase evolved *in vitro* and an existing hepatitis delta virus ribozymes, and evolved these ribozymes so that they would approach each other in sequence space. Among the variants generated they found an "intersection" sequence that possessed both the ligase and the self-cleaving activity supporting the theory that natural ribozymes can evolve from a common ancestor even if they had different structural and functional characteristics.

In vitro selection or directed evolution of nucleic acids involves many of the same steps that are operant during natural selection. *In vitro* selection of aptamers are DNA or RNA molecules that have been selected from random pools based on their ability to bind other molecules [9, 10]. We have previously selected for aptamers that bound to the protein lysozyme, and found that a single binding sequence largely dominated the population [11]. A small number of other species were also isolated, and their relation to the dominant sequence is shown in **Figure 4.1**. The sequences were compared to each other using their pairwise Hamming distances. Hamming distance is a method of sequence comparison that quantifies the dissimilarity between two oligomers by tallying up the number of nucleotide mismatches in their optimal alignment [12]. The dimensionality of the pairwise comparisons where then reduced using Principal Component Analysis. It is possible that this binding sequence represented a 'global optimum' for the selection. However, because the selection experiment generated a fixed

clone, it was unclear whether the evolutionary landscape that connected nucleic acid sequences with lysozyme-binding function contained numerous, discrete local optima; or whether the landscape was similar to the 'Mount Fujiyama' selection function described by Kauffman [13]. More data is essential to populate the sequence space and determine the structure for functional sequence fitness landscapes by investigating how these sequences pan out during *in vitro* selections. Greater number of selection experiments is essential to elucidate the evolutionary pathways that the anti-lysozyme aptamers take in order to attain their final functionality. By probing sequence space with more data, we will be able to populate the fitness landscape providing us with the first glimpse into an evolutionary landscape using real experimental approach. Further analysis of the individual isolated sequences will be tested for their affinity towards the selected target and compared to previously isolated aptamers. The results can shed light into future selection process by taking into account the different factors, which can prevent the isolation of a better aptamer that might have been overlooked. With the advent of automated selection, it is possible to conduct theoretical studies utilizing experimental data to attain a more accurate model.

As a first step towards experimentally probing the nature of evolutionary landscapes using the wet-lab approach, we started from several arbitrary sequences and selected for lysozyme-binding function. A random sequence population of nucleic acids was initially generated by a combination of chemical synthesis and enzymatic amplification. Traditional *in vitro* selection techniques would require weeks to months for one set of selections to be completed, hence, conducting the described selection experiments would prove to be unfeasible. To overcome this obstacle, a Biomek 2000

Laboratory Automation Workstation that has been outfitted was employed to carry out multiple *in vitro* nucleic acid selections in parallel. On the workstation, this population was parsed for function. In this instance, nucleic acids that can bind to an arbitrary target protein, lysozyme, were selected. After each round of selection, the selected nucleic acid binding species (also known as aptamers) were amplified by a combination of reverse transcription, PCR, and *in vitro* transcription. Automated *in vitro* selections allowed the generation of considerable amounts of experimental data in a short amount of time making the link between theoretical studies with real bench work data attainable.

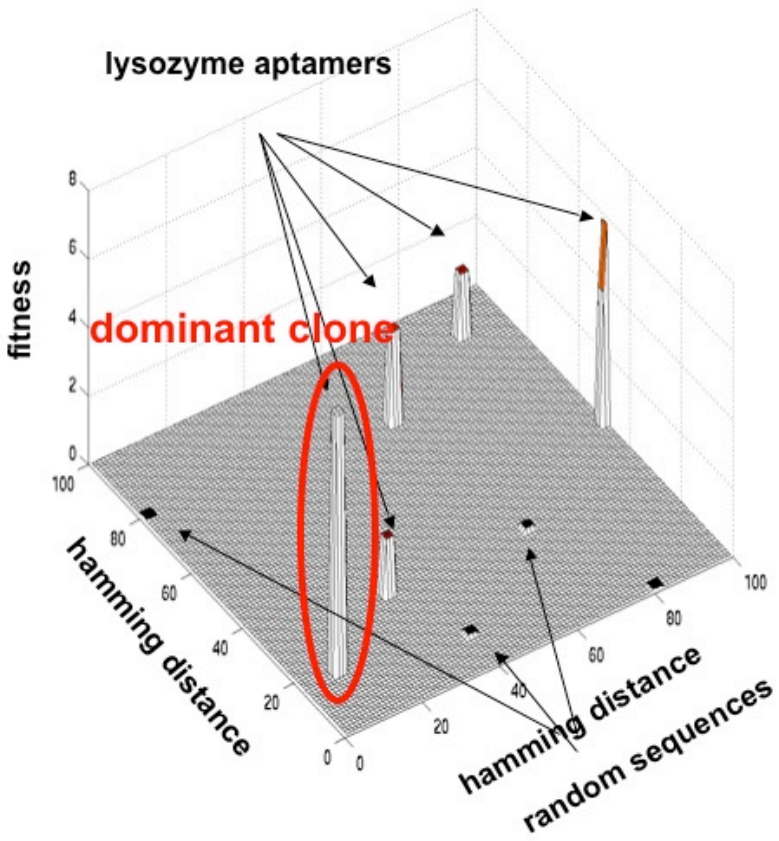


Figure 4.1. Aptamer sequences from the N30 pool selected against lysozyme. The isolated sequences were compared to each other based on their pairwise hamming distances. A principal component analysis was conducted for all the pairwise

comparisons to reduce the data's dimensionality. The 2-dimensional projection of the PCA analysis is plotted on the x and z axis using Matlab. The vertical (y) axis is each aptamer's fitness measure on the binding affinity of the species to the target lysozyme. During this individual selection one dominant clone was isolated. The dominant clone is highlighted in red and shows the highest level of fitness value as measured by their binding affinity to the target lysozyme [14].

In vitro selections are often carried out only once without partitioning the initial pool. The isolated binding aptamer is assumed to be the fittest among the population because it is the sole aptamer isolated. In this experiment, it can be inferred that when given a chance to evolve, other species can be isolated. The results show that no two pools display the same behavior in terms of their evolutionary pattern from their binding trends. During *in vitro* selection, many different factors can prevent the isolation of the best binder. Some RNA aptamers that have better amplification capability might out-compete other more stable structures that are unable to amplify as efficiently. By splitting the pools before the process of selection, we are allowing more room for other species to win in the competition shedding light into whether the previously selected winner is indeed the global optimal aptamer. This work is the progression from work found in the previous section (Chapter 3). Having overcome the contamination issues, we are confident in carrying out theoretical analysis of the data generated with more confidence.

Our approach of first dividing the pool to eight different groups each with unique sequences without any overlap allows us to explore the evolution of aptamer fitness from different starting points in sequence space and a greater number of starting different sequence population. By exploring 8 different pools, we are also able to access a greater number of the pool's population than conducting the selecting using one pool alone. This approach allows each sequence a greater chance to evolve its ability to bind to the target

of interest by giving them a greater chance to be found within each pool population. **Figure 4.2A and Figure 4.2B** compare traditional *in vitro* selections versus our own approach for exploring the aptamer sequence space. It can be discerned from the figures that by exploring evolutionary pathways based on a single pool selection, the results from the experiment can be biased and limited, since only limited amount of the population is accessible to the pool (**Figure 4.2A).** However, when a larger number of pools is tested, more sequences become available to undergo selection. This is a more comprehensive way for populating the aptamer space.

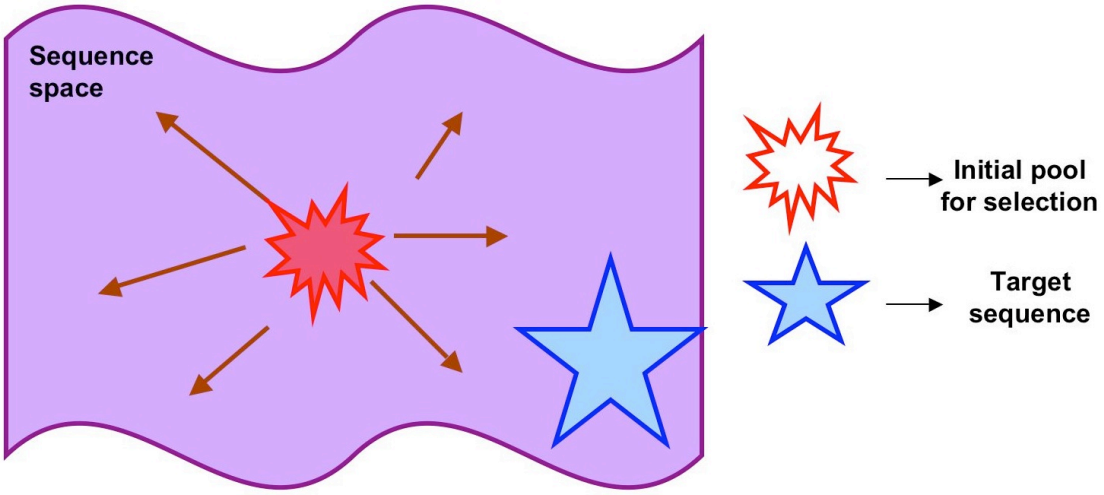


Figure 4.2A. Traditional *in vitro* Aptamer Selections. Selections that are carried out usually explore and start from one single amplified pool and probe for function from that pool. The disadvantage of this approach is that only a small area of the sequence space is being investigated leading to very limited evolutionary pathways these sequences can travel.

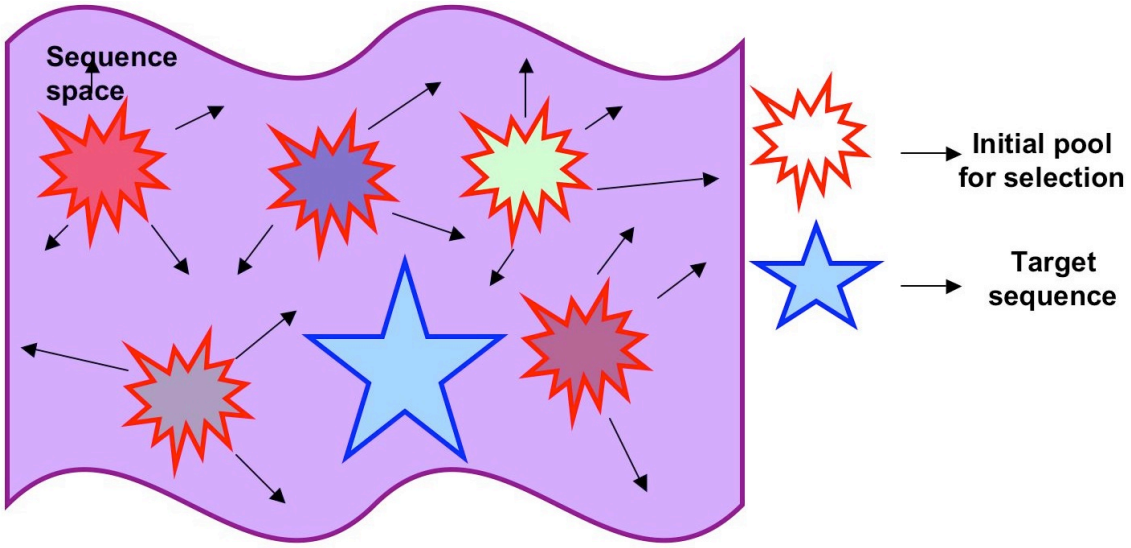


Figure 4.2B. *In vitro* Selection from a Divided Pool. When the initial pool is divided before amplification is introduced, we are able to explore a wider area in the sequence space. This is a more comprehensive way at exploring aptamer sequence space.

RESULTS AND DISCUSSION

For each pool that was harvested from Chapter 3 selections against protein target Lysozyme, sequencing of rounds 6, 12, and 18 were conducted. Individual clones were chosen from the last round of selection (round 18) from the sequence data and assayed for fitness (affinity towards lysozyme). All sequences were assayed in parallel with the previous selected lysozyme aptamer "Clone 1" [11].

POOL 1 FUNCTIONALITY ASSESSMENT

Individual RNA clones were chosen from the round 18 sequences isolated previously as described in Chapter 3. **Figure 4.3A** shows the alignment of round 18 sequences for Pool 1. The sequences chosen for subsequent fitness testing are highlighted in yellow. The approach taken at electing the sequences for testing was based on their frequency of occurrence. If there was a predominant clone, it was crucial to assess their

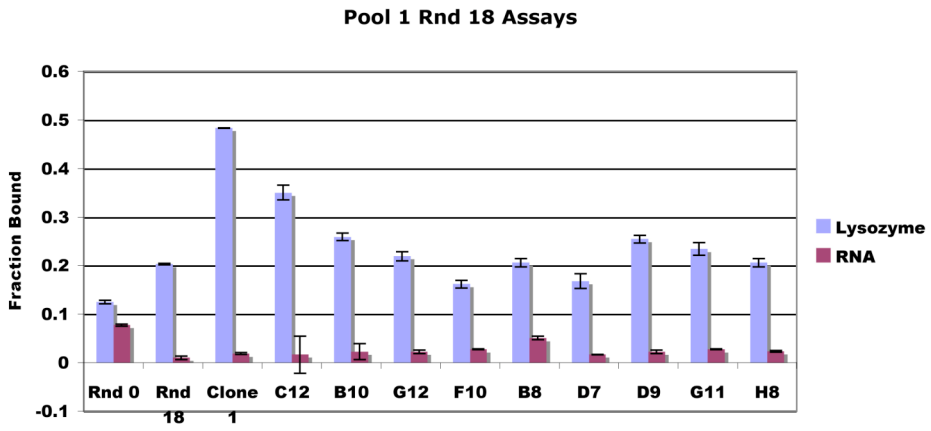
binding fitness and compare to that of previously selected aptamer carried out from one single pool.

Round 18 sequences

```
>C-12GGGAGCATTGCCCACTCATTTTCATTC-AAAAAGCCCGACCCCCATACCCGCACCGACGGTAGCTGCC-GTCACTAGTTCGCGTTGCTG
>G-8 GGGAGCATTGCCCACTCATTTTCATTC-AAAAAGCCCGACCCCCATACCCGAACCGTCCGTAGCTGCC-GTCACTAGTTCGCGTTGCTG

>B-10GGGAGCATTGCCCACTCATTTTCATTC-GGCCGGGGGAAACATGCACAGTAGTTTTGAAGGGGGTTGCG-GTCACTAGTTCGCGTTGCTG
>C-9 GGGAGCATTGCCCACTCATTTTCATTC-GGCCGGGGGAAACATGCACAGTAGTTTTGAAGGGGGTTGCG-GTCACTAGTTCGCGT-GCTG

>H-9 GGGAGCATTGCCCACTCATTTTCATTC-ATTCGCAATAGCAAGGGACGATGCAGAGGATGTCGGGGTG-GTCACTAGTTCGCGTTGCTG
>F-7 GGGAGCATTGCCCACTCATTTTCATTC-ATTCGCAATAGCAAGGGACGATGCAGAGGATGTCGGGGTG-GTCACTAGTTCGCGTTGCTG
>D-9 GGGAGCATTGCCCACTCATTTTCATTC-ATTCGCAATAGCAAGGGACGATGCAGAGGATGTCGGGGTG-GTCACTAGTTCGCGTTGCTG

>G-12GGGAGCATTGCCCACTCATTTTCATTC-GGGGAGCCAGGAATAGAGTACGTCTTAAAAAGAGGCGGGG-GTCACTAGTTCGCGTTGCTG
>F-9 GGGACCATGGCCCATTCATTT-CATTC-GGTGAGCCAAGATTGGAGCGGCGAGACAGAGTAAGTGGCG-GTCACTAGTTCGCGTTGCTG
>E-12GGGAGCATTGCCCA-TCATTTTCATTC-GGAGGCGAAGGGCACTGGCAGATGGACATTCTCGGCCTGG-GTCACTAGTTCGCGTTGCTG
>G-11GGGAGCATTGCCCACTCATTTTCATTC-CCTAGACGCCCATAGATACGCCCCCATGTTGCGCTCGGCC-GTCACTAGTTCGCGTTGCTG
>D-8 GGGAGCATTGCCCACTCATTTTCATTC-CCATTCAATCAGTTAAAACGTTCCCTGGCAATTGGCGGTG-GTCACTAGTTCGCGTTGCTG
>D-7 GGGAGCATTGCCCACTCATTTTCATTC-ACAGTTAAGAAGCATGTTTATCGTGGGTACTCGGGGGTCG-GTCACTAGTTCGCGTTGCTG
>F-8 GGGAGCATTGCCCACTCATTTTCATTC-ATCCGCGAATCCTGAACCCTACGCCATCTGGTCCCGCCCC-GTCACTAGTTCGCGTTGCTG
>C-10GGGAGCATTGCCCACTCATTTTCATTC-CCTAAAACAACATGCTGCACACTGGAAAACCGTCACCCCC-GTCACTAGTTCGCGTTGCTG
>H-8 GGGAGCATTGCCCACTCATTTTCATTC-GAACCCAAGGTGAGTACTCCCCCCCCACTCGGGCCCCCAGG-GTCACTAGTTCGCGTTGCTG
>H-7 GGGAGCATTGCCCACTCATTTTCATTC-CCCCCAGCGGCAGAAGAAAGTTTGCACAAGGTTCACGGTG-GTCACTAGTTCGCGTTGCTG
>C-8 GGGAGCATTGCCCACTCATTTTCATTC-CCCCCGTCGAAACCTCAGCGTAATCCATCTTCTCGCCCC--GTCACTAGTTCGCGTTGCTG
>D-11GGGAGCATTGCCCACTCATTTTCATTC-ACGGAGAGAGCATCGATTATGGGGCCAAACGGGGAGGCTG-GTCACTAGTTCGCGTTGCTG
>F-11GGGAGCATTGCCCACTCATTTTCATTC-GAGCACAGCTATGAAGGCAATTGGCAGGGGTTATTGGCGG-GTCACTAGTTCGCGTTGCTG
>F-10GGGAGCATTGCCCACTCATTTTCATTC-CTGAGGTCCGGGGGATTACGCAGGTACGGGACTAAATGTGC-GTCACTAGTTCGCGTTGCTG
>B-11GGGAGCATTGCCCACTCATTTTCATTC-CTACGCTACCACCGATTGAACCATCCAAAAGAGCATGGAT-GTCACTAGTTCGCGTTGCTG
>B-8 GGGAGCATTGCCCACTCATTTTCATTC-CCCGCCGCATCCGATTTCGCACGACTGAACCGCATTGTCC-GTCACTAGTTCGCGTTGCTG
```

Figure 4.3A. Alignment of Pool 1 Round 18 Sequences. Clone highlighted in yellow are the chosen ones to undergo a binding assay. Primer binding regions are in maroon and flanked the random region of N40. Sequence families are grouped together and colored random region designates identical sequences.

The assays involved the incubation of radioactively transcribed RNA clones with the target protein at equimolar ratios (250 nM) in 100 uL of selection buffer. The reactions were then washed through a vacuum manifold sandwiching a nitrocellulose and a nylon membrane. All reactions were performed in triplicates so that standard errors could be obtained. **Figure 4.3B** shows the binding data for Pool 1 Round 18 selection. The clones isolated from Pool 1 were no better than the previously selected clone 1.




```
>(P1R18)C-12   GGGAGCATTGCCCACTCATTTTCATTC-AAAAAGCCCGACCCCCATACCCGCACCGACGGTAGCTGCC-GTCACTAGTTCGCGTTGCTG
>(P1R18)B-10   GGGAGCATTGCCCACTCATTTTCATTC-GGCCGGGGAAACATGCACAGTAGTTTTGAAGGGGGTTGCG-GTCACTAGTTCGCGTTGCTG
>(P1R18)G-12   GGGAGCATTGCCCACTCATTTTCATTC-GGGGAGCCAGGAATAGAGTACGTCTTAAAAAGAGGCGGGG-GTCACTAGTTCGCGTTGCTG
>(P1R18)F-10   GGGAGCATTGCCCACTCATTTTCATTC-CTGAGGTCCGGGGATTACGCAGGTACGGGACTAAATGTGC-GTCACTAGTTCGCGTTGCTG
>(P1R18)B-8    GGGAGCATTGCCCACTCATTTTCATTC-CCCGCCGCATCCGATTTCGCACGACTGAACCGCATTGTCC-GTCACTAGTTCGCGTTGCTG
>(P1R18)D-7    GGGAGCATTGCCCACTCATTTTCATTC-ACAGTTAAGAAGCATGTTTATCGTGGGTACTCGGGGGTCG-GTCACTAGTTCGCGTTGCTG
>(P1R18)D-9    GGGAGCATTGCCCACTCATTTTCATTC-ATTCGCAATAGCAAGGGACGATGCAGAGGATGTCGGGGTG-GTCACTAGTTCGCGTTGCTG
>(P1R18)G-11   GGGAGCATTGCCCACTCATTTTCATTC-CCTAGACGCCCATAGATACGCCCCCATGTTGCGCTCGGCC-GTCACTAGTTCGCGTTGCTG
>(P1R18)H-8    GGGAGCATTGCCCACTCATTTTCATTC-GAACCCAAGGTGAGTACTCCCCCCCCACTCGGGCCCCCAGG-GTCACTAGTTCGCGTTGCTG
>Clone 1       GGGAATGGATCCACATCTACGAATTC---ATCAGGGCTAAAGAGTGCAGAGTTACTTAG----TTCACTGCAGACTTGACGAAGCTT
```

Figure 4.3B. Binding Data for Individual Clones from Pool 1 at Round 18.

All clones, except for C12, had affinities to the target comparable to that of Round 18 assayed as a pool. This data is not surprising considering the information about the drop in binding affinity that occurred when the selection went from round 12 to 18. This selection did not reach a level of pool enrichment that is often observed during *in vitro* selections of aptamers.

POOL 2 FUNCTIONALITY ASSESSMENT

Since Pool 2 did not show any predominant sequence or common motifs, sequences were chosen at random. However, the presence of a drop in binding from

round 12 to round 18 made investigating clones in round 12 important. From the sequence alignments shown in **Figure 4.4A** of both round 12 and 18, there was a hint of species enrichment that was starting to take place, but the sequence clone (B4) disappeared by the time round 18 was reached.

Looking at the binding affinities **(Figure 4.4B)**, it showed relatively high fitness for both round 12 clones. Not only were the numbers comparable to that of round 18, B2 had an affinity equal to Clone 1 previously selected from the N30 pool. Even with such high affinity towards the protein target, these clones disappeared from the population at round 12. One would expect clone B2 to evolve and overtake the entire selection because of its high binding to Lysozyme as expected from traditional selection methods. The disappearance of such high binding molecule could be attributed to the limits of amplification during *in vitro* selections as discussed in Chapter 3. This data provided further evidence that stable, high performance sequences might be at a disadvantage during replication [15].

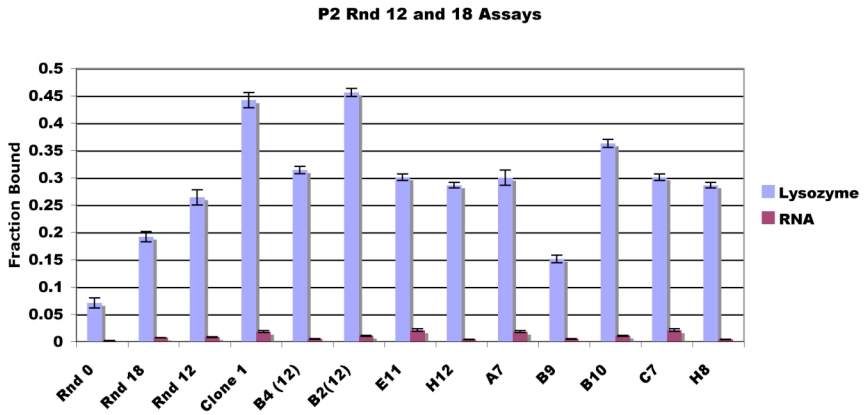
Round 12 sequences

```
>A2  GGGAGCATTGCCCACTCATTTTCATTC-TGACAAGTGCAACCGTACACCGGTTTATGCTGGTAGAATA-GTCACTAGTTCGCGTTGCTG
>B4  GGGAGCATTGCCCACTCATTTTCATTC-TGACAAGTGCAACCGTACACCGGTTTATGCTGGTAGAATA-GTCACTAGTTCGCGTTGCTG

>G3  GGGAGCATTGCCCACTCATTTTCATTC-GCTCGTCGATTAGGGTCATAGCAATCATACTTTAATTAAG-GTCACTAGTTCGCGTTGCTG
>H3  GGGAGCATTGCCCACTCATTTTCATTC-GCTCGTCGATTAGGGTCATAGCAATCATACTTTAATTAAG-GTCACTAGTTCGCGTTGCTG
>D3  GGGAGCATTGCCCACTCATTTTCATTC-GACGTTTATTAACGGAATAGCCTACTGGTGCAACTACTGC-GTCACTAGTTCGCGTTGCTG
>F4  GGGAGCATTGCCCACTCATTTTCATTC-TTGTTAGGGTACTTTCTGTCAACGTGACATAGCTTTATTA-GTCACTAGTTCGCGTTGCTG
>B2  GGGAGCATTGCCCACTCATTTTCATTC-TATGTCCACAAACGACACTTCACACGATTTAAATTCCGCT-GTCACTAGTTCGCGTTGCTG
>C3  GGGAGCATTGCCCACTCATTTTCATTC-ACCGCACTCCAGAGGCTACGGGTTAAATTTCCTGACATGG-GTCACTAGTTCGCGTTGCTG
```

Round 18 sequences

```
>E9   GGGAGCATTGCCCACTCATTTTCATTC-GGAAGTGTAGGAGAACAGTAAGCTTAGAAACAATGGCCGGGGTCACTAGTTCGCGTTGCTG
>F10  GGGAGCATTGCCCACTCATTTTCATTC-GGAGGACAAAGTGGAACAGACTGTACAACAAGGCCGGC---GTCACTAGTTCGCGTTGCTG
>C10  GGGAGCATTGCCCACTCATTTTCATTC-GGCTTCGTACGGCGGGGTCCATAGGAAGATATCTCAGTG--GTCACTAGTTCGCGTTGCTG
>A7   GGGAGCATTGCCCACTCATTTTCATTC-CCCCCGACACTACCCCCGCAGCTACAACCAGGTCCCGGTC-GTCACTAGTTCGCGTTGCTG
>H8   GGGAGCATTGCCCACTCATTTTCATTC-CCCCCGCCGAACAGCTGCAGAAAGATAGAGCTCCTCCGCC-GTCACTAGTTCGCGTTGCTG
>C11  GGGAGCATTGCCCACTCATTTTCATTC-CCCCCTCGCAGGATGTAGAACTACAAAAACTCGGACCGCC-GTCACTAGTTCGCGTTGCTG
>C7   GGGAGCATTGCCCACTCATTTTCAT-C-CCCCCGCATAGGTACTAATACGAAAGCAAGATCGACCCGCCGTCACTAGTTCGCGTTGCTG
>H11  GGGAGCATTGCCCACTCATTTTCATTC-CCCTAGGTGTGCTATACCCCGGAAAGAACTACGACCCCCC-GTCACTAGTTCGCGTTGCT-
>G9   GGGAGCATTGCCCACTCATTTTCATTC-CCCTCAACCGTGGCATAGGTAGGCTCAATGGACGGACGTG-GTCACTAGTTCGCGTTGCTG
>B9   GGGAGCATTGCCCACTCATTTTCATTC-CCTCCCGTGAAGAGTACGAACTAACAAACCTACCCCTGGC-GTCACTAGTTCGCGTTGCTG
>G12  GGGAGCATTGCCCACTCATTTTCATTC-CCGAACTGAGAACAGTCCGTAACGACAACCCCCACAGCCC-GTCACTAGTTCGCGTTGCTG
>H9   GGGAGCATTGCCCACTCATTTTCATTC-GTCAAAGCAAAGAGCTGACAGATAAGAACCAACCCCGGCC-GTCACTAGTTCGCGTTGCTG
>D9   GGGAGCATTGCCCTTTCATTTTCATTC-CCGGAATACAGAGAAGTAGTAACAAAGATTTGACCCGGCC-GTCACTAGTTCGCGTTGCTG
>B7   GGGAGCATTGCCCACTCATTTTCATTC-CCGTACACGCACCAACGACCAGCAGAGATACAACCCGGCC-GTCACTAGTTCGCGTTGCTG
>C8   GGGAGCATTGCCCACTCATTTTCATTT-CTTGTCTGACGCCGGCCAGTAGTCCGTCATGCACCCCGTGCGTCACTAGTTCGCGTTGCTG
>G10  GGGAGCATTGCCCACTCATTTTCATTC-CACCCTCCGAACTATACAAAACAAGACCGACCCGTCCTGC-GTCACTAGTTCGCGTTGCTG
>E11  GGGAGCATTGCCCACTCATTTTCATTC-AAGGATAGCGGGAGTAATCGTCGAGCCATGTTGTACGGGG-GTCACTAGTTCGCGTTGCTG
>E12  GGGAGCATTGCCCACTCATTTTCATTC-AGGCGGAGGACAAGACAGCATATGGGGACCCCCGGGTATC-GTCACTAGTTCGCGTTGCTG
>B11  GGGAGCATTGCCCACTCATTTTCATTC-ACAACAGAAACCCCCGATCCGAACCATTACCCCGACCCG--GTCACTAGTTCGCGTTGCTG
>H10  GGGAGCATTGCCCACTCATTTTCATTC-AGACCGAACAAGAACGCACATAATACCAAACCCCCCCGGACGTCACTAGTTCGCGTTGCTG
>D12  GGGAGCATTGCCCACTCATTTTCATTC-AGCAGCCCACCTCCGATCCCTCATACCTAACCCGCCGGTC-GTCACTAGTTCGCGTTGCTG
>E10  GGGAGCATTGCCCACTCATTTTCATTC-ACACCGAAGAACAGGATCATAGTACAAGCACACCGGGGCC-GTCACTAAGCCGAATTCCAG
>H12  GGGAGCATTGCCCACTCATTTTCATTC-ATCCCCAAAGGAACGGAGGGAAGAAGTCGATACAGGCCGG-GTCACTAGTTCGCGTTGCTG
>F9   GGGAGCATTGCCCACTCATTTTCATTC-ATGAGACCCCCTGCATTCCAACGAATGGAACCCCCCTGC--GTCACTAGTTCGCGTTGCT-
>H7   GGGAGCATTGCCCACTCATTTTCATTC-AATGGGGACGGAGAACAGAATCAACGATTGGCAGTTGGGG-GTCACTAGTTCGCGTTGCTG
>F7   GGGAGCATTGCCCACTCATTTTCATTC-AGGGAAGGCAGGGGAGAACCAAGAGGAACTTTGGCGCGTTGGTCACTAGTTCGCGTTGCTG
```

Figure 4.4A. Alignment of Pool 2 Round 12 and 18 Sequences. Clones highlighted in yellow are the chosen ones to undergo a binding assay. Primer binding regions are in maroon and flanked the random region of N40. Sequence families are grouped together and colored random region designates identical sequences.




```
>(P2R12)B-4   GGGAGCATTGCCCACTCATTTTCATTC-TGACAAGTGCAACCGTACACCGGTTTATGCTGGTAGAATA-GTCACTAGTTCGCGTTGCTG
>(P2R12)B-2   GGGAGCATTGCCCACTCATTTTCATTC-TATGTCCACAAACGACACTTCACACGATTTAAATTCCGCT--GTCACTAGTTCGCGTTGCTG
>(P2R18)E-11  GGGAGCATTGCCCACTCATTTTCATTC-AAGGATAGCGGGAGTAATCGTCGAGCCATGTTGTACGGGG--GTCACTAGTTCGCGTTGCTG
>(P2R18)H-12  GGGAGCATTGCCCACTCATTTTCATTC-GATGCAAGCTACGGACAATCAGAAGACACCGGCGCCCATG--GTCACTAGTTCGCGTTGCTG
>(P2R18)A-7   GGGAGCATTGCCCACTCATTTTCATTC-CCCCCGACACTACCCCCGCAGCTACAACCAGGTCCCGGTC--GTCACTAGTTCGCGTTGCTG
>(P2R18)B-9   GGGAGCATTGCCCACTCATTTTCATTC-CCTCCCGTGAAGAGTACGAACTAACAAACCTACCCCTGGC--GTCACTAGTTCGCGTTGCTG
>(P2R18)C-7   GGGAGCATTGCCCACTCATTTTCAT-C-CCCCCGCATAGGTACTAATACGAAAGCAAGATCGACCCGCC-GTCACTAGTTCGCGTTGCTG
>(P2R18)H-8   GGGAGCATTGCCCACTCATTTTCATTC-CCCCCGCCGAACAGCTGCAGAAAGATAGAGCTCCTCCGCC--GTCACTAGTTCGCGTTGCTG
>Clone 1   GGGAATGGATCCACATCTACGAATTC------ATCAGGGCTAAAGAGTGCAGAGTTACTTAG-------TTCACTGCAGACTTGACGAAGCTT
```

Figure 4.4B. Binding Data for Individual Clones from Pool 2 at Round 12 and Round 18.

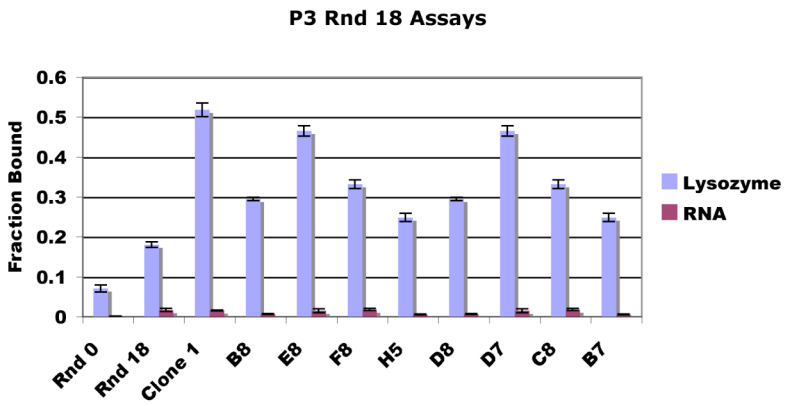
POOL 3 FUNCTIONALITY ASSESSMENT

Pool 3 sequence enrichment trends were very typical for that of a traditional selection (**Figure 4.5A)**. Round 18 was predominantly populated by one single clone sequence. This dominant RNA was tested along with other sequence clones. Two of the isolated clones showed high binding, however, neither one of them were the dominant sequence (**Figure 4.5B**). This supports the notion that the frequency of a clone does not necessary mean that it has optimal functionality. Moreover, during *in vitro* selections many factors, in addition to just function, plays an important role in the deciding the species will come out being the winner.

Round 18 sequences

```
>F7   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>F6   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>D5   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>C6   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>H7   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>H6   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>H5   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCCAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>B5   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>F5   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>E7   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>B6   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGTATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>A7   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>C7   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>G7   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>A8   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>E6   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>E5   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>D8   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGAAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>A5   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>G8   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGAAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>C5   GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCAGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>G5   GGGAGCATTGCCCACTCATTTTCATTC-CAAACAACGTAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG

>B7   GGGAGCATTGCCCACTCATTTTCATTC-TCAACGCACAATGCAAGTCAGCGACATGCTATGCCGGCCC-GTCACTAGTTCGCGTTGCTG
>F8   GGGAGCATTGCCCACTCATTTTCATTC-CAAAAGCGACAATGGTTTCGGGGTCAGCTGAGCGTGTCGG-GTCACTAGTTCGCGTTGCTG
>B8   GGGAGCATTGCCCACTCATTTTCATTC-TCCAAAAATCCACGCGACATACACAAAACTTGCTGGGCCG-GTCACTAGTTCGCGTTGCTG
>C8   GGGAGCATTGCCCACTCATTTTCATTC-CGAGACCCAGAGCACTCCAGTCCTGCATTCCCCACTGGTG-GTCACTAGTTCGCGTTGCTG
>D7   GGGAGCATTGCCCACTCATTTTCATTC-GCCAATGCGTTTTCAGGACTATAATGGACCAACAGGGGGC-GTCACTAGTTCGCGTTGCTG
>E8   GGGAGCATTGCCCACTCATTTTCATTC-ACGAACGTCATCCTTAGCCGCACGCCGTCCGGCATTCCGC-GTCACTAGTTCGCGTTGCTG
```

Figure 4.5A. Alignment of Pool 3 Round 18 Sequences. Clones highlighted in yellow are the chosen ones to undergo a binding assay. Primer binding regions are in maroon and flanked the random region of N40. Sequence families are grouped together and colored random region designates identical sequences.




```
>(P3R18)B-8  GGGAGCATTGCCCACTCATTTTCATTC-TCCAAAAATCCACGCGACATACACAAAACTTGCTGGGCCG-GTCACTAGTTCGCGTTGCTG
>(P3R18)E-8  GGGAGCATTGCCCACTCATTTTCATCC-CAAATAACGTAATACTCGGAATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>(P3R18)F-8  GGGAGCATTGCCCACTCATTTTCATTC-CAAAAGCGACAATGGTTTCGGGGTCAGCTGAGCGTGTCGG-GTCACTAGTTCGCGTTGCTG
>(P3R18)H-5  GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGTAATACTCGGCATTTCCAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>(P3R18)D-8  GGGAGCATTGCCCACTCATTTTCATTC-CAAATAACGAAATACTCGGCATTTCAAAACCCCTTCGGAC-GTCACTAGTTCGCGTTGCTG
>(P3R18)D-7  GGGAGCATTGCCCACTCATTTTCATTC-GCCAATGCGTTTTCAGGACTATAATGGACCAACAGGGGGC-GTCACTAGTTCGCGTTGCTG
>(P3R18)C-8  GGGAGCATTGCCCACTCATTTTCATTC-CGAGACCCAGAGCACTCCAGTCCTGCATTCCCCACTGGTG-GTCACTAGTTCGCGTTGCTG
>(P3R18)B-7  GGGAGCATTGCCCACTCATTTTCATTC-TCAACGCACAATGCAAGTCAGCGACATGCTATGCCGGCCC-GTCACTAGTTCGCGTTGCTG

>Clone 1  GGGAATGGATCCACATCTACGAATTC------ATCAGGGCTAAAGAGTGCAGAGTTACTTAG-------TTCACTGCAGACTTGACGAAGCTT
```

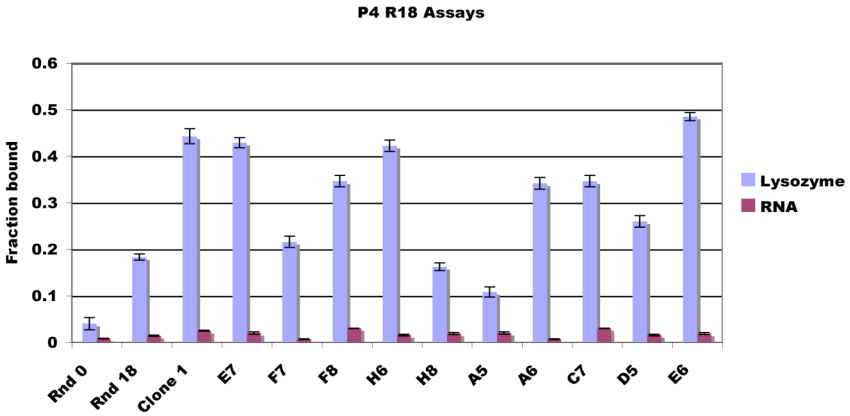
Figure 4.5B. Binding Data for Individual Clones from Pool 3 at Round 18.

POOL 4 FUNCTIONALITY ASSESSMENT

At round 18 for Pool 4, there was a slight level of enrichment for a number of isolated individuals **(Figure 4.6A)**. Although the pool itself did not show groundbreaking binding, the clones isolated; however, show binders that were comparable to Clone 1 from the N30 pool selection. We were starting to see the ability to isolate aptamers to Lysozyme that showed affinities similar or even higher to that of Clone 1 by splitting the initial pool eight different ways (**Figure 4.6B**).

```
Round 18 sequences

>C7   GGGAGCATTGCCCACTCATTTTCATTC-TCCGAGCATGGATCTACCGTGGACTACCATCGTTATCCTC-TCACTAGTTCGCGTTGCTG
>D6   GGGAGCATTGCCCACTCATTTTCATTC-TCCGAGCGGAGTCTGCTTGATTCTTTACCAAATCTATTAT-TCACTAGTTCGCGTTGCTG

>G5   GGGAGCATTGCCCACTCATTTTCATTC-AAATAACACCTGATTCAGTTTTAATCTTGTAATCGTGGT--TCACTAGTTCGCGTTGCTG
>E6   GGGAGCATTGCCCACTCATTTTCATTC-AAATAACACCTGATTCAGTTTTAATCTTGTAATCGTGGT--TCACTAGTTCGCGTTGCTG

>F8   GGGAGCATTGCCCACTCATTTTCATTC-TGTCCGGACTAACGCCAGTACGAGATTTTTCCGAGCCAGG-TCACTAGTTCGCGTTGCTA
>A6   GGGAGCATTGCCCACTCATTTTCATTC-AGCTACAGTTGCTGTGGAAATTTTCCGATGTGTGCGTCCA-TCACTAGTTCGCGTTGCTA

>E7   GGGAGCATTGCCCACTCATTTTCATTC-ATAAAATAAAATTGTAATCACGGTGCTCGAACTATGGCGTGTCACTAGTTCGCGTTGCTG
>H6   GGGAGCATTGCCCACTCATTTTCATTC-TTAGCACCAGCTAAAACACATAAAATATCCCTTGTGGCTG-TCACTAGTTCGCGTTGCTG

>D5   GGGAGCATTGCCCACTCATTTTCATTC-AGCGGCGCACTTGAGCGCAAGATCAGCAGGACTTCCGAGT-TCACTAGTTCGCGTTGCTA
>F7   GGGAGCATTGCCCACTCATTTTCATTC-CCGATCAGCGGCAGTTGGTTTCGCCGTGCCGTTCACAGA--TCACTAGTTCGCGTTGCTG
>A5   GGGAGCATTGCCCACTCATTTTCATTC-AGGGATTGCCCCTTCTCAGAGCGGTGCGCTCCTTTGGAGG-TCACTAGTTCGCGTTGCTG
>A8   GGGAGCATTGCCCACTCATTTTCATTC-TGCATCTCATCAGATGCTTGATGAGAAGCGCCTGTTCATC-TCACTAGTTCGCGTTGCTG
>G8   GGGAGCATTGCCCACTCATTTTCATTC-AATTTGTGTAAGTTAATCTTAGCTCCGGTGTAAGTCTGAG-TCACTAGTTCGCGTTGCTG
>H7   GGGAGCATTGCCCACTCATTTTCATTC-TCCTAACAAACTGTCCGATATTAACGGAACACCTATCCGA-TCACTAGTTCGCGTTACTA
>G6   GGGAGCATTGCCCACTCATTTTCATTC-ACGCAAACTCCAAGTCTCCGCAACACAGGATACCCCTTAC-TCACTAGTTCGCGTTGCTA
>C8   GGGAGCATTGCCCACTCATTTTCATTC-CGAACCAAGTTAAAATCTCGCACCCAGAGTGGAGATACC--TCACTAGTTCGCGTTGCTG
>E8   GGGAGCATTGCCCACTCATTTTCATTC-AATGTAGGTCCACATGTGCAAACTATAGTCAGTCAAAGGT-TCACTAGTTCGCGTTGCTG
>H8   GGGAGCATTGCCCACTCATTTTCATTC-TACCGCGGAATAGAGTGTCCAACGGCTTCCGTTGGGGCTG-TCACTAGTTCGCGTTGCTG
>C6   GGGAGCATTGCCCACTCATTTTCATTC-TAAGCTGGTATGCTTAGAATTCATGCCTTCTTGTAACGCC-TCACTAGTTCGCGTTACTG
>G7   GGGAGCATTGCCCACTCATTTTCATTC-GGCGTCCGGTTAGTGATTGGTAATGCCGGACGGTGGTGTA-TCACTAGTTCGCGTTGCTG
>B6   GGGAGCATTGCCCACTCATTTTCATTC-GGTTGGATCTAATGACGCCAGGAGTGACGGGCACTCGATG-TCACTAGTTCGCGTTGCTG
>A7   GGGAGCATTGCCCACTCATTTTCATTC-TGATTGCTAGTTACTAGCAGAATAAACCTTGTCTGATGC--TCACTAGTTCGCGTTACTA
```

Figure 4.6A. Alignment of Pool 4 Round 18 Sequences. Clones highlighted in yellow are the chosen ones to undergo a binding assay. Primer binding regions are in maroon and flanked the random region of N40. Sequence families are grouped together and colored random region designates identical sequences.




```
>(P4R18)E-7   GGGAGCATTGCCCACTCATTTTCATTC-ATAAAATAAAATTGTAATCACGGTGCTCGAACTATGGCGT-GTCACTAGTTCGCGTTGCTG
>(P4R18)F-7   GGGAGCATTGCCCACTCATTTTCATTC-CCGATCAGCGGCAGTTGGTTTCGCCGTGCCGTTCACAGA--GTCACTAGTTCGCGTTGCTG
>(P4R18)F-8   GGGAGCATTGCCCACTCATTTTCATTC-TGTCCGGACTAACGCCAGTACGAGATTTTTCCGAGCCAGG-GTCACTAGTTCGCGTTGCTG
>(P4R18)H-6   GGGAGCATTGCCCACTCATTTTCATTC-TTAGCACCAGCTAAAACACATAAAATATCCCTTGTGGCTG-GTCACTAGTTCGCGTTGCTG
>(P4R18)H-8   GGGAGCATTGCCCACTCATTTTCATTC-TACCGCGGAATAGAGTGTCCAACGGCTTCCGTTGGGGCTG-GTCACTAGTTCGCGTTGCTG
>(P4R18)A-5   GGGAGCATTGCCCACTCATTTTCATTC-AGGGATTGCCCCTTCTCAGAGCGGTGCGCTCCTTTGGAGG-GTCACTAGTTCGCGTTGCTG
>(P4R18)A-6   GGGAGCATTGCCCACTCATTTTCATTC-AGCTACAGTTGCTGTGGAAATTTTCCGATGTGTGCGTCCA-GTCACTAGTTCGCGTTGCT-
>(P4R18)C-7   GGGAGCATTGCCCACTCATTTTCATTC-TCCGAGCATGGATCTACCGTGGACTACCATCGTTATCCTC-GTCACTAGTTCGCGTTGCTG
>(P4R18)D-5   GGGAGCATTGCCCACTCATTTTCATTC-AGCGGCGCACTTGAGCGCAAGATCAGCAGGACTTCCGAGT-GTCACTAGTTCGCGTTGCT-
>(P4R18)E-6   GGGAGCATTGCCCACTCATTTTCATTC-AAATAACACCTGATTCAGTTTTAATCTTGTAATCGTGGT--GTCACTAGTTCGCGTTGCT-

>Clone 1   GGGAATGGATCCACATCTACGAATTC------ATCAGGGCTAAAGAGTGCAGAGTTACTTAG-------TTCACTGCAGACTTGACGAAGCTT
```

Figure 4.6B. Binding Data for Individual Clones from Pool 4 at Round 18.

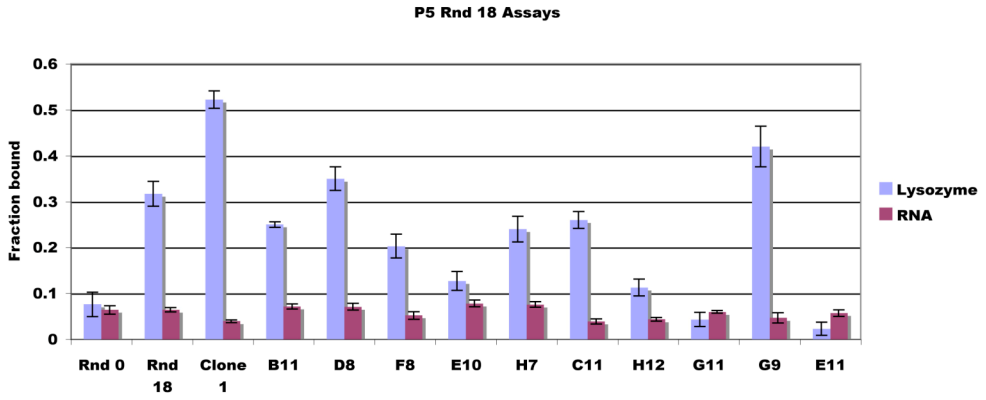
POOL 5 FUNCTIONALITY ASSESSMENT

Since there was no predominant family or sequence clone that was found in round 18 in Pool 5 clones are chosen at random for assaying as shown in **Figure 4.7A** The pool itself bound at a high fraction. However the individual clones were not as impressive as one would have expected. A couple of clones showed binding fractions almost as high as that of Clone 1 (**Figure 4.7B**). Since no clear loss of complexity is observed for this pool, under normal circumstances, the selection of this pool would continue beyond round 18 with increases in stringency.

Round 18 sequences

```
>F8   GGGAGCATTGCCCACTCATTTTCATTC-CCCCTGCGATGGAAGAGGAAACAGACATGAACTGGCCCGG-GTCACTAGTTCGCGTTGCTG
>C12  GGGAGCATTGCCCACTCATTTTCATTC-ACGAACAGAACCAGCCTCACGACCCCTCCAAACCCCGGCC-GTCACTAGTTCGCGTTGCTG
>G12  GGGAGCATTGCCCACTCATTTTCATTC-CGCGACCCTATTGAACACCCCGCCCTTGGAACCCCCGGCC-GTCACTAGTTCGCGTTGCTG
>E10  GGGAGCATTGCCCACTCATTTTCATTC-GACCCGTCGCAGACCCCATCAGAGCTACTGATCGCCCCCG-GTCACTAGTTCGCGTTGCTG
>H7   GGGAGCATTGCCCACTCATTTTCATTC-GACCCGAAGCCCAAAGAACCTAGAGAGCAACCACCCGGCC-GTCACTAG-TCGCGTTGCTG
>C11  GGGAGCATTGCCCACTCATTTTCATTC-GCGAAACCAACCAAGCCCCGTATGGAAGATAGACCCGGCC-GTCACTAGTTCGCGTTGCTG
>H12  GGGAGCATTGCCCACTCATTTTCATTC-ACGTAGCGAGAGAACGATAAGTACAGAGAACACGGCCCGG-GTCACTAGTTCGCGTTGCTG
>G11  GGGAGCATTGCCCACTCATTTTCATTC-GCTGAGAGGGTGCAGGTCTACTGAATGGGCTAGCCGTGGG-GTCACTAGTTCGCGTTGCTG
>G9   GGGAGCATTGCCCACTCATTTTCATTC-GAAGCGAAGATAAAGCAAGAAGCACTATCCAGGCCGGCC--GTCACTAGTTCGCGTTGCTG
>G7   GGGAGCATTGCCCACTCATTTTCATTC-GGTGGGAGGACTACGTCAAGCAACTACATACATGGGGCCG-GTCACTAGTTCGCGTTGCTG
>E11  GGGAGCATTGCCCACTCATTTTCATTC-GAACAGTAGAGAAAGCAGAAGACGTAAACGAAGCGCCGGG-GTCACTAGTTCGCGTTGCTG
>F12  GGGAGCATTGCCCACTCATTTTCATTC-AAGGACGGGGGCAGAACACGCATCAGAACTTAACGGCCGG-GTCACTAGTTCGCGTTGCTG
>E7   GGGAGCATTGCCCACTCATTTTCATTC-GCAGGGGGAAGCTTGTAGAACAATGGTACCGACGCCGGGG-GTCACTAGTTCGCGTTGCTG
>H9   GGGAGCATTGCCCACTCATTTTCATTC-ACAGAACCGACATGACCCCCGCAGCTACAAGACCGCCGCC-GTCACTAGTTCGCGTTGCTG
>D8   GGGAGCATTGCCCACTCATTTTCATTC-AGGGAAGGCCACAATAGACGAACACAAGTATCCGCGGGG--GTCACTAGTTCGCGTTGCTG
>E12  GGGAGCATTGCCCACTCATTTTCATTC-GAAACCAGAAAATATCCCAACGTCAATCAGACACGGCCGG-GTCACTAGTTCGCGTTGCTG
>B11  GGGAGCATTGCCCACTCATTTTCATTC-CCACAAAAAGGCAAAGAGTATCGAGAAGGTCACCCCCGGC-GTCACTAGTTCGCGTTGCTG
>A9   GGGAGCATTGCCCACTCATTTTCATTC-GCCAGCGTTTGCCACAGACACCACAACGACATCGGCCCGG-GTCACTAGTTCGCGTTGCTG
>A10  GGGAGCATTGCCCACTCATTTTCATTC-AGCACAGAGTACGTAAGGATAAGGAATACAACTGCCGGGG-GTCACTAGTTCGCGTTGCTG
>F11  GGGAGCATTGCCCACTCATTTTCATTC-CAGGAAGCTAGAAGGGAGGAACAAACAGCTTGAACCGCCC-GTCACTAGTTCGCGTTGCTG
>B9   GGGAGCATTGCCCACTCATTTTCATTC-GCAGGATGGGTAAGGCAAATTGGACAGTTGATTAGCGGGG-GTCACTAGTTCGCGTTGCTG
```

Figure 4.7A. Alignment of Pool 5 Round 18 Sequences. Clones highlighted in yellow are the chosen ones to undergo a binding assay. Primer binding regions are in maroon and flanked the random region of N40. Sequence families are grouped together and colored random region designates identical sequences.




```
> (P5R18)B-11 GGGAGCATTGCCCACTCATTTTCATTC-CCACAAAAAGGCAAAGAGTATCGAGAAGGTCACCCCCGGC-GTCACTAGTTCGCGTTGCTG
> (P5R18)D-8  GGGAGCATTGCCCACTCATTTTCATTC-AGGGAAGGCCACAATAGACGAACACAAGTATCCGCGGGG--GTCACTAGTTCGCGTTGCTG
> (P5R18)F-8  GGGAGCATTGCCCACTCATTTTCATTC-CCCCTGCGATGGAAGAGGAAACAGACATGAACTGGCCCGG-GTCACTAGTTCGCGTTGCTG
> (P5R18)E-10 GGGAGCATTGCCCACTCATTTTCATTC-GACCCGTCGCAGACCCCATCAGAGCTACTGATCGCCCCCGCCGTCACTAGTTCGCGTTGCTG
> (P5R18)H-7  GGGAGCATTGCCCACTCATTTTCATTC-GACCCGAAGCCCAAAGAACCTAGAGAGCAACCACCCGGCC--GTCACTAGTTCGCGTTGCTG
> (P5R18)C-11 GGGAGCATTGCCCACTCATTTTCATTC-GCGAAACCAACCAAGCCCCGTATGGAAGATAGACCCGGCC--GTCACTAGTTCGCGTTGC-G
> (P5R18)H-12 GGGAGCATTGCCCACTCATTTTCATTC-ACGTAGCGAGAGAACGATAAGTACAGAGAACACGGCCCGG--GTCACTAGTTCGCGTTG-TG
> (P5R18)G-11 GGGAGCATTGCCCACTCATTTTCATTC-GCTGAGAGGGTGCAGGTCTACTGAATGGGCTAGCCGTGGG--GTCACTAGTTCGCGTT-CTG
> (P5R18)G-9  GGGAGCATTGCCCACTCATTTTCATTC-GAAGCGAAGATAAAGCAAGAAGCACTATCCAGGCCGGCC---GTCACTAGTTCGCGTG-CTG
> (P5R18)E-11 GGGAGCATTGCCCACTCATTTTCATTC-GAACAGTAGAGAAAGCAGAAGACGTAAACGAAGCGCCGGG--GTCACTAGTTCGCGTTGCTG

>Clone 1   GGGAATGGATCCACATCTACGAATTC------ATCAGGGCTAAAGAGTGCAGAGTTACTTAG-------TTCACTGCAGACTTGACGAAGCTT
```

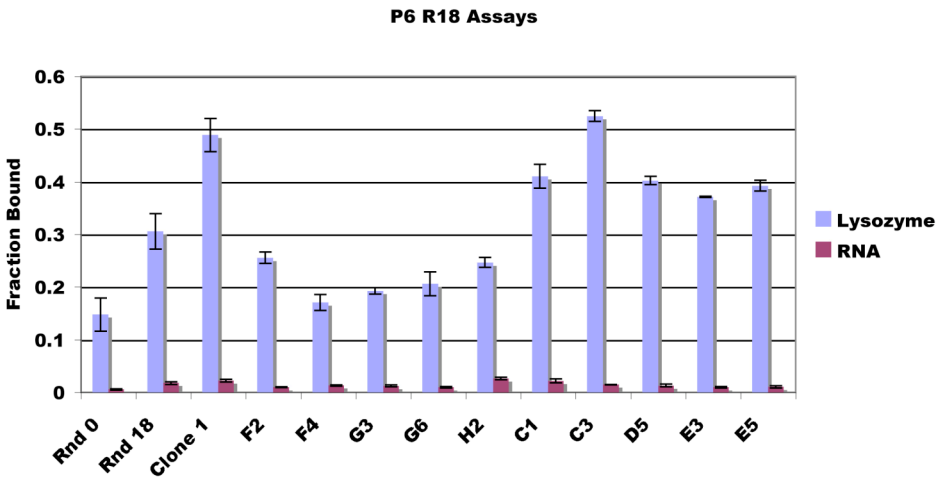
Figure 4.7B. Binding Data for Individual Clones from Pool 5 at Round 18.

POOL 6 FUNCTIONALITY ASSESSMENT

Pool 6 showed a high level of enrichment. However, multiple species were becoming more frequent in the population at round 18 (**Figure 4.8A).** The binding data showed that not all of the species exhibited high degree of binding (**Figure 4.8B).** The majority of the clones bound at the same level as the pool itself only. This is another example of loss in complexity does not necessary translate to the isolation of a good binder.

```
Round 18 sequences

>E5 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>E6 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>E4 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>E2 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>C5 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>C2 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>B2 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>H6 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAGCATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>A6 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG
>H1 GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG--GTCACTAGTTCGCGTTGCTG

>H2 GGGAGCATTGCCCACTCATTTTCATAC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGG---GTCACTAGTTCGCGTTGCTG
>G1 GGGAGCATTGCCCACTCATTTTCATTC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGGG--GTCACTAGTTCGCGTTGCTG
>D6 GGGAGCATTGCCCACTCATTTTCATTC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGGG--GTCACTAGTTCGCGTTGCTG
>A2 GGGAGCATTGCCCACTCATTTTCATTC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGG---GTCACTAGTTCGCGTTGCTG
>B3 GGGAGCATTGCCCACTCATTTTCATTC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGGG--GTCACTAGTTCGCGTTGCAG
>F1 GGGAGCATTGCCCACTCATTTTCATTC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGGG--GTCACTAGTTCGCGTTGCTG


>F2 GGGAGCATTGCCCACTCATTTTCATTC-GAGCAGCAAACTGGAAGGGATAAGATAAGAAGTCGCCGGG--GTCACTAGTTCGCGTTGCTG
>H3 GGGAGCATTGCCCACTCATTTTCATTC-GAGCAGCAAACTGGAAGGGATAAGATAAGAAGTCGCCGGG--GTCACTAGTTCGCGTTGCTG
>E1 GGGAGCATTGCCCACTCATTTTCATTC-GAGCAGCAAACTGGAAGGGATAAGATAAGAAGTCGCCGGG--GTCACTAGTTCGCGTTGCTG
>A3 GGGAGCATTGCCCACTCATTTTCGTTC-GAGCAGCAAACTGGAAGGGATAAGATAAGAAGTCGCCGGG--GTCACTAGTTCGCGTTGCTG

>E3 GGGAGCATTGCCCACTCATTTTCATTC-GACCCCCCGGATAATCGACCCGCAGAGCCTAGACCCCGCC--GTCACTAGTTCGCGTTGCTG
>B6 -GGAGCATTGCCCACTCATTTTCATTC-GACCCCCCGGATAATCGACCCGCAGAGCCTAGACCCCGCC--GTCACTAGTTCGCGTTGCTG
>G5 GGGAGCATTGCCCACTCATTTTCATTC-GACCCCCCGGATAACCGACCCGCAGAGCCTAGACCCCGCC--GTCACTAGTTCGCGTTGCTG

>F4 GGGAGCATTGCCCACTCATTTTCATTC-GAAGCGTCAGCAGTCACGGAATACATTGTCGGTCGGGGTG--GTCACTAGTTCGCGTTGCTG
>H4 GGGAGCATTGCCCACTCATTTTCATTC-GAAGCGTCAGCAGTCACGGAATACATTGTCGGTCGGGGTG--GTCACTAGTTCGCGTTGCTG

>G6 GGGAGCATTGCCCACTCATTTTCA--C-CGAACCATACACCTTGACCACCCCCCAAGAGAGCCCCGCC--GTCACTAGTTCGCGTTGCTG
>A4 GGGAGCATTGCCCACTCATTTTCA--C-CGAACCATACACCTTGACCACCCCCCAAGAGAGCCCCGCC--GTCACTAGTTCGCGTTGCTG

>G3 GGGAGCATTGCCCACTCATTTTCATTC-CCTCCACGAGTATAAGAACAAAAGACAGAAACCCCCGGGCCCGTCACTAGTTCGCGTTGCTG
>B1 -GGAGCATTGCCCACTCATTTTCATTC-CCTCCACGAGTATAAGAACAAAAGACAGAAACCCCCGGGCCCGTCACTAGTTCGCGTTGCTG

>D5 GGGAGCATTGCCCACTCATTTTCATTC-CCAACCCCGCTAAGAACAAGAAAGCCACGAGTCCCCGGCC--GTCACTAGTTCGCGTTGCTG
>G2 GGGAGCATTGCCCACTCATTTTCATTC-CCAACCCCGCTCAGAACAAGAAAGCCACGAGTCCCCGGCC--GTCACTAGTTCGCGTTGCTG

>C1 GGGAGCATTGCCCACTCATTTTCATTC-AACAGCGGATAGAACGTAGCTATAAGACAGTTTGGCGGGG--GTCACTAGTTCGCGTTGCTG
>D1 GGGAGCATTGCCCACTCATTTTCATTC-AACAGCGGATAGAACGTAGCTATAAGACAGTTTGGCGGGG--GTCACTAGTTCGCGTTGCTG
>C6 GGGAGCATTGCCCACTCATTTTCATTC-AACGGATAGAGCATAAAGTCGAAGCATCAACTTGGCGGG---GTCACTAGTTCGCGTTGCTG

>C3 GGGAGCATTGCCCACTCATTTTCATCC-GAACGATAACGAAGATACGCCCTAGATTTCCCCGCCGCC---GTCACTAGTTCGCGTTGCTG
>F3 GGGAGCATTGCCCACTCATTTTCATTC-GAATGAAGGCTAAACGACAAGGAAAGACATTTCGGCGGGG--GTCACTAGTTCGCGTTGCTG

>A5 GGGAGCATTGCCCACTCATTTTCATTC-GAAGAGGAGAGTAAACAGGGACGAACACCGTACTGGCCGG--GTCACTAGTTCGCGTTGCTG
>A1 -GGAGCATTGCCCACTCATTTTCATTC-CCGCGATACCTGCCCCAACGAACGAGACCACAGACCCGCC--GTCACTAGTTCGCGTTGCTA
>F5 GGGAGCATTGCCCACTCATTTTCATTC-GCAACGACAACATCGCTTGAGATACATACATGCCACGGGG--GTCACTAGTTCGCGTTGCTG
>H5 GGGAGCATTGCCCACTCATTTTCATTC-GCGCCAGAAATCGCCCGACCCAGCACCACTTTGGACCGCC--GTCACTAGTTCGCGTTGCTG
```

Figure 4.8A. Alignment of Pool 6 Round 18 Sequences. Clones highlighted in yellow are the chosen ones to undergo a binding assay. Primer binding regions are in maroon and flanked the random region of N40. Sequence families are grouped together and colored random region designates identical sequences.




```
>P6R18)F-2  GGGAGCATTGCCCACTCATTTTCATTC-GAGCAGCAAACTGGAAGGGATAAGATAAGAAGTCGCCGGG---GTCACTAGTTCGCGTTGCTG
>P6R18)F-4  GGGAGCATTGCCCACTCATTTTCATTC-GAAGCGTCAGCAGTCACGGAATACATTGTCGGTCGGGGTG---GTCACTAGTTCGCGTTGCTG
>P6R18)G-3  GGGAGCATTGCCCACTCATTTTCATTC-CCTCCACGAGTATAAGAACAAAAGACAGAAACCCCCGGGCCC-GTCACTAGTTCGCGTTGCTG
>P6R18)G-6  GGGAGCATTGCCCACTCATTTTCA--C-CGAACCATACACCTTGACCACCCCCCAAGAGAGCCCCGCC---GTCACTAGTTCGCGTTGCTG
>P6R18)H-2  GGGAGCATTGCCCACTCATTTTCATAC-GCCACAAAGAAGGGGTAAAGTCACGAAACGAACGGCCGG----GTCACTAGTTCGCGTTGCTG
>P6R18)C-1  GGGAGCATTGCCCACTCATTTTCATTC-AACAGCGGATAGAACGTAGCTATAAGACAGTTTGGCGGGG---GTCACTAGTTCGCGTTGCTG
>P6R18)C-3  GGGAGCATTGCCCACTCATTTTCATCC-GAACGATAACGAAGATACGCCCTAGATTTCCCCGCCGCC----GTCACTAGTTCGCGTTGCTG
>P6R18)D-5  GGGAGCATTGCCCACTCATTTTCATTC-CCAACCCCGCTAAGAACAAGAAAGCCACGAGTCCCCGGCC---GTCACTAGTTCGCGTTGCTG
>P6R18)E-3  GGGAGCATTGCCCACTCATTTTCATTC-GACCCCCCGGATAATCGACCCGCAGAGCCTAGACCCCGCC---GTCACTAGTTCGCGTTGCTG
>P6R18)E-5  GGGAGCATTGCCCACTCATTTTCATTC-ACACGAGCTACAGCATAAGACACAGGGAAACATCGGCCGG---GTCACTAGTTCGCGTTGCTG

>Clone 1   GGGAATGGATCCACATCTACGAATTC------ATCAGGGCTAAAGAGTGCAGAGTTACTTAG-------TTCACTGCAGACTTGACGAAGCTT
```

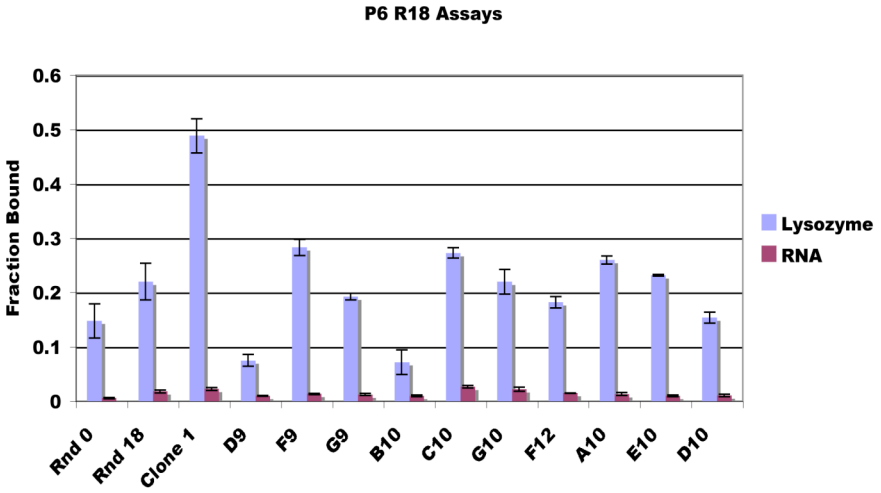
Figure 4.8B. Binding Data for Individual Clones from Pool 6 at Round 18.

POOL 7 FUNCTIONALITY ASSESSMENT

Pool 7 remained diverse even at round 18 (**Figure 4.9A).** Not only was the pool still highly complex, the biding of Pool 7 at round 18 was only slightly above background (**Figure 4.9B**), which is reflected on the isolated clones as well. They all exhibited mediocre binding percentages. This pool is a great candidate to undergo further rounds of selection with increasing stringency needed.

```
Round 18 sequences

>D9  GGGAGCATTGCCCACTCATTTTCATTC-GCAGAATAAAGCCCCGTAATACGACCTTTGGACCCGCC---GTCACTAGTTCGCGTTGCTG
>F9  GGGAGCATTGCCCACTCATTTTCATTC-GAAGAGCACTGGGGATGAATCAGCATGTTTGTGCACCGGG-GTCACTAGTTCGCGTTGCTG
>G9  GGGAGCATTGCCCACTCATTTTCATTC-GACAAAGAATAACAAAGGCACAAGAGCACAGTATGGGGGG-GTCACTAGTTCGCGTTGCTG
>B10 GGGAGCATTGCCCACTCATTTTCATTC-TGCAAACACAGACCCCCCGGAAGGCGATCAAAACACCCCGCGTCACTAGTTCGCGTTGCTG
>C10 GGGAGCATTGCCCACTCATTTTCATTC-ACCAAGAAAAGAGGAGAAAACACAGGGAACTATCGGCG---GTCACTAGTTCGCGTTGCTG
>G10 GGGAGCATTGCCCACTCATTTTCATTC-AGGGAGCACAATGGAGTCTACAAGCGTAAGTCTGGCCGGG-GTCACTAGTTCGCGTTGCTG
>F12 GGGAGCATTGCCCACTCATTTTCATTC-GACCCCCAACAACTACGAATATGTACAACGACCGGCCCTT-GTCACTAGTTCGCGTTGCTG
>A10 GGGAGCATTGCCCACTCATTTTCAT-C-CCCCGCCGAAAGGTAGAGAAAACACGACCAGAGACCGCCC-GTCACTAGTTCGCGTTGCTG
>E10 GGGAGCATTGCCCACTCATTTTCATTC-CGAGGCAAGGGAAGTTAGAGGTCAAGAGATTAGCGCTGGG-GTCACTAGTTCGCGTTGCTG
>D10 GGGAGCATTGCCCACTCATTTTCATTC-GAAGGACGAGCAGGGTATACACCAAATGGAACACGGCGGG-GTCACTAGTTCGCGTTGCTG
>E11 GGGAGCATTGCCCACTCATTTTCATTC-AGGGCAGGAAGAAACTGGTAGGATGCACTTGACGCCGGGG-GTCACTAGTTCGCGTTGCTG
>F10 GGGAGCATTGCCCACTCATTTTCATTC-GAGTTGGGGAAGGGCAGTCACATAGAATAAGGTCGGGCC--GTCACTAGTTCGCGTTGCTG
>H11 GGGAGCATTGCCCACTCATTTTCATTC-GGAGGGGCTAAGGACACAAGAGCAAACACTTGGCATGGGG-GTCACTAGTTCGCGTTGCTG
>H10 GGGAGCATTGCCCACTCATTTTCATTC-GGGGGGACAACGTAAACATGAGGCAAGGATCGCAGTTCAC-GTCACTAGTTCGCGTTGCTG
>E12 GGGAGCATTGCCCACTCATTTTCATTC-CCGTAGGCAGCTAGAACCAAAAGTCGAGAGTACCCGGCCC-GTCACTAGTTCGCGTTGCTG
```

Figure 4.9A. Alignment of Pool 7 Round 18 Sequences. Clones highlighted in yellow are the chosen ones to undergo a binding assay. Primer binding regions are in maroon and flanked the random region of N40. Sequence families are grouped together and colored random region designates identical sequences.




```
>P7R18-D9  GGGAGCATTGCCCACTCATTTTCATTC-GCAGAATAAAGCCCCGTAATACGACCTTTGGACCCGCC---GTCACTAGTTCGCGTTGCTG
>P7R18F9   GGGAGCATTGCCCACTCATTTTCATTC-GAAGAGCACTGGGGATGAATCAGCATGTTTGTGCACCGGG-GTCACTAGTTCGCGTTGCTG
>P7R18G9   GGGAGCATTGCCCACTCATTTTCATTC-GACAAAGAATAACAAAGGCACAAGAGCACAGTATGGGGGG-GTCACTAGTTCGCGTTGCTG
>P7R18B10  GGGAGCATTGCCCACTCATTTTCATTC-TGCAAACACAGACCCCCCGGAAGGCGATCAAAACACCCCGCGTCACTAGTTCGCGTTGCTG
>P7R18C10  GGGAGCATTGCCCACTCATTTTCATTC-ACCAAGAAAAGAGGAGAAAACACAGGGAACTATCGGCG---GTCACTAGTTCGCGTTGCTG
>P7R18G10  GGGAGCATTGCCCACTCATTTTCATTC-AGGGAGCACAATGGAGTCTACAAGCGTAAGTCTGGCCGGG-GTCACTAGTTCGCGTTGCTG
>P7R18F12  GGGAGCATTGCCCACTCATTTTCATTC-GACCCCCAACAACTACGAATATGTACAACGACCGGCCCTT-GTCACTAGTTCGCGTTGCTG
>P7R18A10  GGGAGCATTGCCCACTCATTTTCAT-C-CCCCGCCGAAAGGTAGAGAAAACACGACCAGAGACCGCCC-GTCACTAGTTCGCGTTGCTG
>P7R18E10  GGGAGCATTGCCCACTCATTTTCATTC-CGAGGCAAGGGAAGTTAGAGGTCAAGAGATTAGCGCTGGG-GTCACTAGTTCGCGTTGCTG
>P7R18D10  GGGAGCATTGCCCACTCATTTTCATTC-GAAGGACGAGCAGGGTATACACCAAATGGAACACGGCGGG-GTCACTAGTTCGCGTTGCTG


>Clone 1   GGGAATGGATCCACATCTACGAATTC------ATCAGGGCTAAAGAGTGCAGAGTTACTTAG-------TTCACTGCAGACTTGACGAAGCTT
```

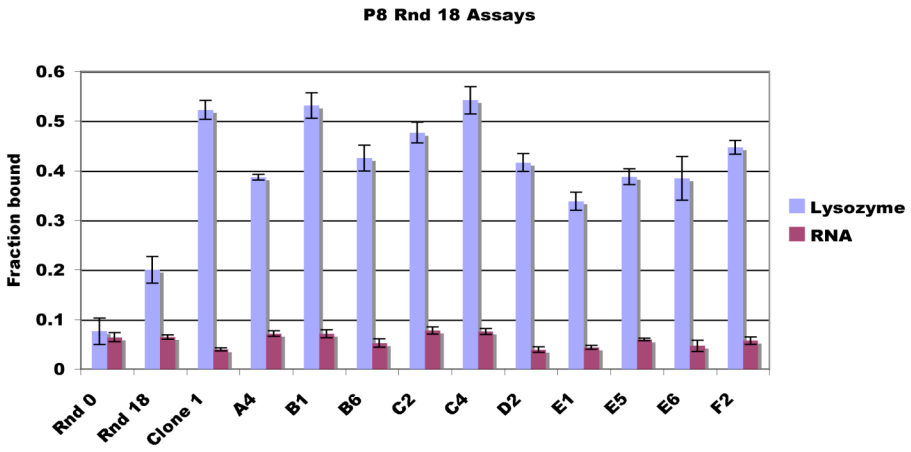
Figure 4.9B. Binding Data for Individual Clones from Pool 7 at Round 18.

POOL 8 FUNCTIONALITY ASSESSMENT

Very high degree of enrichment occurred with Pool 8 as seen in **Figure 4.10A.** The isolation of comparable and some better binders to Clone 1 was supported by high degree of complexity reduction (**Figure 4.10B)**. Their bound fractions were very similar across all the clones, giving the impression of a family coexisting and evolving together.

```
Round 18 sequences

>C2  GGGAGCATTGCCCACTCATTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>B2  GGGAGCATTGCCCACTTTTTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>G1  GGGAGCATTGCCCACTCATTT-CATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>E2  GGGAGCATTGCCCACTCTTTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>F4  AGGAGCATTGCCCACTCTTTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAACGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>H6  GGGAGCATTGCCCACTCATTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>A6  GGGAGCATTGCC-ACTCATTTTCATTC-GTCGATAAGGACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>H3  GGGAGCATTGCCCACTCATTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>G6  GGGAGCATTGCCCATTTATTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>F3  GGGAGCATTGCCCACTCATTTTCATTC-GTCGATA-GTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>C1  GGGAGCATTGCCCACTCATTTTCATTC-GTGGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG

>E1  GGGAGCATTGCCCACTCATTTTCATTC-CCAAGGCATTGATGATGGCGATCTCAAGAATATGACGGCGGTG-GTCACTAGTTCGCGTTGCTG
>G5  GGGAGCATTGCCCACTCATTTTCATTC-CCAAGGCATTGATGATGGCGATTAAGAAATGACGGCGGTG----GTCACTAGTTCGCGTTGCTG
>G2  GGGAGCATTGCCCACTCTTTTTCATTC-CCAAGGCATTGATGATGGCGATTAAGGTATGACGGCGGTG----GTCACTAGTTCGCGTTGCTG
>H4  GGGAGCATTGCCCACCTCATTTCATTC-CCAAGGCATTGATGATGGCGATTAAGGTATGACGGCGGTG----GTCACTAGTTCGCGTTGCTG
>H2  GGGAGCATTGCCCACTCATTTTCATTC-CCAAGGCATTGATGATGGCGATTAAGGTATGACGGCGGTG----GTCACTAGTTCGCGTTGCTG

>B6  GGGAGCATTGCCCATTTATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTTGACACCGGCCCCC---GTCACTAGTTCGCGTTGCTG
>B4  GGGAGCATTGCCCATTCATGTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTCGACACCGGCCGCC----GTCACTAGTTCGCGT-GCTG
>E3  GGGAGCATTGCCCTTTCATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTCGACACCGGCCCCC----GTCACTAGTTCGCGTTGCTG
>D1  -GGAGCATTGCCCACTCATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTCGACACCGGCCCCC----GTCACTAGTTCGCGTTGCTG
>A3  GGGAGCATTGCCCATTTATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTCGACACCGGCCCCC----GTCACTAGTTCGCGTTGCTG
>F1  GGGAGCATTGCCCATTCATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTCGACACCGGCCCCC----GTCACTAGTTCGCGTTGCTG
>C5  GGGAGCATTGCCCACTCATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTACGACACCGGCCCCC---GTCACTAGTTCGCGTTGCTG

>E5  GGGAGCATTGCCCACTCATTTTCATTC-AGCGTAAATGGGCACTTACGAAATAACAAGGTCGGTGAGG----GTCACTAGTTCGCGTTGCTG
>G3  GGGAGCATTGCCCACTCATTTTCATTC-AGCGTAAATGGGCACTTACGAAATAACACGGTCGGTGAGG----GTCACTAGTTCGCGTTGCTG
>D3  GGGAGCATTGCCCACTCATTTTCATTC-AGCGTAAATGGGCACTTACGAAATAACACGGTCGGTGTGG----GTCACTAGTTCGCGTTGCTG

>B1  GGGAGCATTGCCCACTCATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCTG
>B3  GGGAGCATTGCCCATTCATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCTG
>G4  GGGAGCATTGCCCATTCATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCTG
>C3  GGGAGCATTGCCCATTCATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCAG
>H1  GGGAGCATTGCCCACTCATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCTG

>C4  GGGAGCATTGCCCATTTATTTTCATTC-AGTACCTGACCAGCACCCGTAGGGATCCTGACCCGCCGCC----GTCACTAGTTCGCGTTGCTG
>E4  GGGAGCATTGCCCTTTCATTTTCATTC-AGTACCTGACCAGCACCCGTAGGGATCCTGACCCGCCGCC----GTCACTAGTTCGCGTTGCTG

>D2  GGGAGCATTGCCCACTCATTTTCATTC-CAGAACCCCCCTCGAGAAAGCTACACCTTAGTCCCCCGGC----GTCACTAGTTCGCGTTGCTG
>D4  GGGAGCATTGCCCACTCATTTTCATTC-CAGACCCCCCCTCGAGAAAGCTACACCTTAGTCCCCCGGC----GTCACTAGTTCGCGTTGCTG

>F2  GGGAGCATTGCCCATCTATTTTCATTC-CACGCACCCAGATGCTAACTGAAACGCACCACCCGCCGGC----GTCACTAGTTCGCGTTGCTG
>A5  GGGAGCATTGCCCACTCATTTTCATTC-CACGCACCCAGATGCTAACTGAAACGCACCACCCGCCGGC----GTCACTAGTTCGCGTTGCTG
>F5  GGGAGCATTGCCCATTCATTTTCACTC-GCATCTGAACAAGGATACGTAGCGTTTAGAACCCCCCGCC----GTCACTAGTTCGCGTT----
>B5  GGGAGCATTGCCCACTCATTTTCATTC-GCACACACAGCAATGATAGTTCATCAAGACGAAGCCGGGG----GTCACTAGTTCGCGTTGCTG
>A4  GGGAGCATTGCCCATTTATTTTCATTC-AGGATGAGAGCGTACATGGTAGAGACAACGAGCCGGCGTG----GTCACTAGTTCGCGTTGCTG
>E6  GGGAGCATTGCCCACTCATTTTCATTC-TAAGAGGGGGGTCCATCGCAAACAGAACACGGATGTCGGCA---GTCACTAGTTCGCGTTGCTG
```

Figure 4.10A. Alignment of Pool 8 Round 18 Sequences. Clones highlighted in yellow are the chosen ones to undergo a binding assay. Primer binding regions are in maroon and flanked the random region of N40. Sequence families are grouped together and colored random region designates identical sequences.




```
>P8R18)A-4 GGGAGCATTGCCCATTTCATTTTCATTC-AGGATGAGAGCGTACATGGTAGAGACAACGAGCCGGCGTG----GTCACTAGTTCGCGTTGCTG
>P8R18)B-1 GGGAGCATTGCCCACTC-ATTTTCAT-C-GGACCCCCTGCGAGAACAGCATCGTCCACGTCGCCCTCAC----GTCACTAGTTCGCGTTGCTG
>P8R18)B-6 GGGAGCATTGCCCATTTCATTTTCATTC-AGCCAAGCCGCCCCTCCGAACTCACTTGACACCGGCCCCC----GTCACTAGTTCGCGTTGCTG
>P8R18)C-2 GGGAGCATTGCCCACTC-ATTTTCATTC-GTCGATAAGTACTCAGCAGGGAATATACAGCGGTTCGGGG----GTCACTAGTTCGCGTTGCTG
>P8R18)C-4 GGGAGCATTGCCCATTTCATTTTCATTC-AGTACCTGACCAGCACCCGTAGGGATCCTGACCCGCCGCC----GTCACTAGTTCGCGTTGCTG
>P8R18)D-2 GGGAGCATTGCCCACTC-ATTTTCATTC-CAGAACCCCCCTCGAGAAAGCTACACCTTAGTCCCCCGGC----GTCACTAGTTCGCGTTGCTG
>P8R18)E-1 GGGAGCATTGCCCACTC-ATTTTCATTC-CCAAGGCATTGATGATGGCGATCTCAAGAATATGACGGCGGTG-GTCACTAGTTCGCGTTGCTG
>P8R18)E-5 GGGAGCATTGCCCACTC-ATTTTCATTC-AGCGTAAATGGGCACTTACGAAATAACAAGGTCGGTGAGG----GTCACTAGTTCGCGTTGCTG
>P8R18)E-6 GGGAGCATTGCCCACTC-ATTTTCATTC-TAAGAGGGGGGTCCATCGCAAACAGAACACGGATGTCGGCA---GTCACTAGTTCGCGTTGCTG
>P8R18)F-2 GGGAGCATTGCCCATCTCATTTTCATTC-CACGCACCCAGATGCTAACTGAAACGCACCACCCGCCGGC----GTCACTAGTTCGCGTTGCTG

>Clone 1   GGGAATGGATCCACATCTACGAATTC------ATCAGGGCTAAAGAGTGCAGAGTTACTTAG-------TTCACTGCAGACTTGACGAAGCTT
```

Figure 4.10B. Binding Data for Individual Clones from Pool 8 at Round 18.

MULTIVARIATE PROJECTION METHODS

The main challenge in genotype space representation is the ability to display data from multivariate datasets. A method known as multivariate projection is widely used for the analysis and modeling of multivariate datasets. This method is used in the analysis of sequence space data presented in this work. The main multivariate projection method

applied in chemometrics is the Principal Component Analysis (PCA). This method is defined as a linear transformation that converts the data to a new coordinate, such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on and so forth. PCA reduces the dimension in a dataset by retaining those properties of the data set that contribute to the most variance also known as the "most important" aspects of the data.

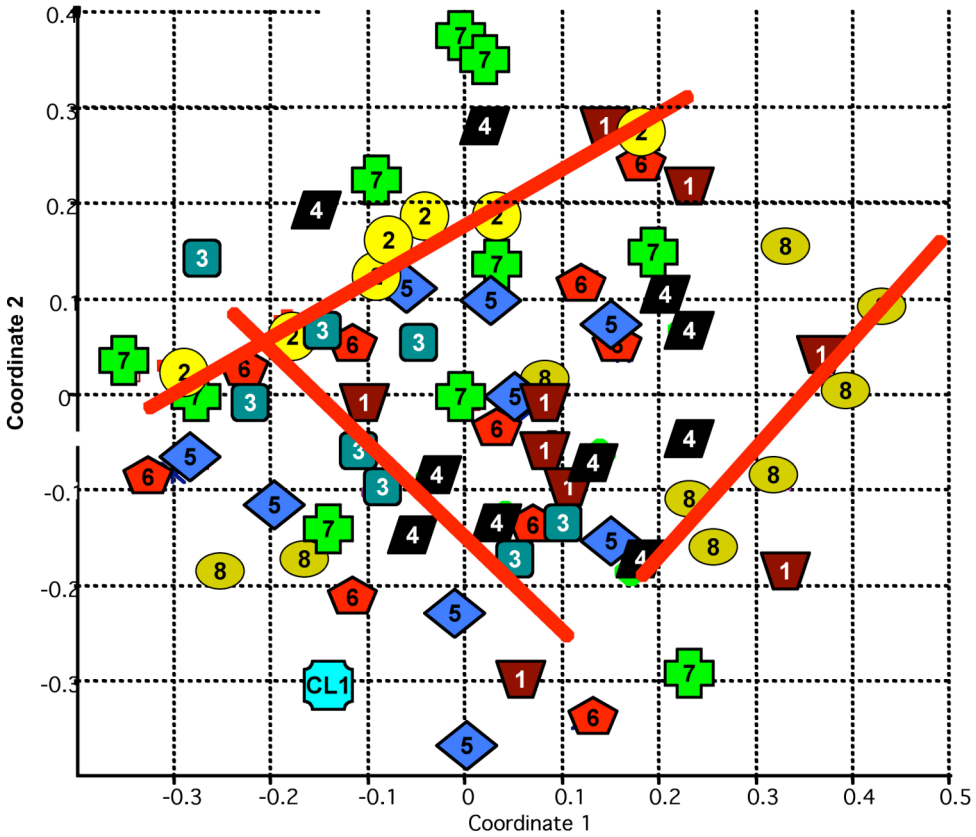
PCA is highly suited for the analysis of data that cannot be conducted using the traditional statistical standards. Common traits with biological datasets are that they contain more variables than observations where the information is noisy. Underlying structures and latent variables are important parts included in these datasets. A latent variable cannot be measured directly but is estimated as a linear combination of directly measured variables. A dataset consisting of N observations and K variables can theoretically be visualized as a plot of the N samples in a K dimensional space. However, the limitations of our cognitive abilities make it very difficult to visualize such spaces when $K > 3$ or more than 3 dimensions. The latent variables can be geometrically interpreted as vectors in the multidimensional space that describe as much as possible of the variation in the data and minimize the sum of the squared residuals. The combination of multiple components creates hyper-planes. Multivariate projection methods are transparent and can be easily interpreted and validated in depth. The dimensional reduction of components from the multidimensional space gives results that can be geometrically interpreted in two-dimensional and three-dimensional plots. Other methods used for modeling multivariate data are artificial neural networks (ANNs) and Support

Vector Machines (SVMs), which are referred to as nonlinear methods. Unlike in SVM and ANN, where data are implicitly transformed in the modeling process, PCA requires the data to be explicitly transformed or expanded with quadratic terms in order to handle nonlinearity.

Figure 4.11 is a PCA analysis of the sequence data of the clones that have been tested for each pool. The analysis is conducted using PCOORD (http://www.hiv.lanl.gov/content/hiv-db/PCOORD/PCOORD.html). In summary, sequence alignments of the clones that have been assayed are submitted. A pairwise hamming distance matrix is generated and by utilizing sequence ordination [16] to obtain PCA values are obtained and plotted allowing 10 dimensions for the user to select in order to obtain a two-dimensional view of this multi-dimensional analysis.

From the PCA analysis, pools selected that have reached some level of functional equilibrium by the appearance of dominant sequences in addition to displaying high degree of target affinity fall within a dimension. For instance, sequences isolated at round 18 from Pool 2 fall almost in a linear function to the red line that is added to help with visualization. Pool 8 is another example of the appearance of several RNA sequence families at round 18. Pool 3 shows similar trend as that of Pool 2 and 8. Clone 1 is also plotted to show its position relative to other functional sequences.

Although all the isolated clones (including the ones from Pools 2,3, and 8) have similar functions with comparable levels of binding capacity, they do not share common primary sequences. They can be as different from each other as some the sequences are from the previously selected Clone 1.



,

Figure 4.11. PCA Analysis of the Round 18 Tested Sequences. Two-dimensional view of the Principal Coordinate Analysis (PCA) is shown. Coordinates 1 and 2 are displayed. As can be inferred from the graph, most sequences are extremely different from each other as well as the previously isolated Clone 1. However, for pools 2, 3, and 8 where strong binders were isolated and families formed, their sequences are more alike within the pool as shown by the linear display of the aptamer layout. In conclusion, the traditional 1 pool 1 selection traditional method does not really allow for the isolation of the best true binder, as this analysis shows, the fitness landscapes still contains other sequences with functionality fitness comparable or even better than the previously isolated sequence.

CONCLUSION

From this analysis, it can be inferred that *in vitro* selections carried out with one single pool limits the window for achieving other equally good binders. Single selections

have limited evolutionary paths as well as nucleic acid species that can be explored. We have isolated aptamers that showed higher or equal degree of activity than was previously thought possible. If only the individual sequences within a population were given a chance to evolve without being constricted within one single pool, more areas of the functional and sequence space could be accessed. The data here shows that the fitness landscape of aptamers or unnatural sequences is not composed of that one single peak but rather multiple peaks. From the data presented here, there are certainly limitations to conventional *in vitro* selections at attaining the best aptamer. One of the biggest constraints is the replication efficiency for reactions conducted *in vitro*. This is another area that the lab has started to explore.

Even in these experiments, we might not have even found the optimal sequence; after all, there are constraints when performing *in vitro* selections that cannot replicate the real events that occur natural selection. In this work, we have confirmed that the straightforward single pool has a lot of limitations in terms of the number of sequences that it can be accessed and the limited number of species with good fitness that can be isolated. As the methods for nucleic acid space continues to be improved by *in vitro* selections, more questions can begin to be addressed such as whether methods to search the entire sequence space for functions can ever be attained. What are the difficulties and properties that hinder the evolution of a molecule for certain functionality *in vitro*? Understanding these aspects will eventually allow the scientists to faithfully replicate nature's approach at generating molecular function.

MATERIALS AND METHODS

INDIVIDUAL CLONE SEQUENCING

Individual clones from each selected pool were isolated and sequenced. Once the products are harvested from the robot selection process, a small aliquot from the RT-PCR reactions for rounds 6, 12, and 18 are amplified and then ligated into a thymidine-overhang vector (TA Cloning Kit, Invitrogen). The templates that will undergo sequencing reactions are generated from colony PCR reactions [17] using standard M13 reverse and forward sequencing primers that flank the insertion site of the TA vector. These PCR products are purified using a 96-well Multiscreen-PCR cleanup plate (Millipore). The integrity and size of the colony PCR products are visualized on a 4% agarose gel to ensure the success of template insertion.

Sequencing reactions are performed using a CEQ DTCS Quick Start Kit (Beckman-Coulter). The reactions are performed as described by the vendor's instructions. In summary, about 100 fmol of purified products are used as templates. The unincorporated dyes are with dry Sephadex G50 (Amersham Pharmacia) placed into a MultiScreen HV plate. The sephadex is hydrated with water for 3 hours at room temperature. The sequencing reactions are then loaded onto the columns and spun for 5 minutes at 1,100 g. The samples recovered are then dried under vacuum and the pellets are resuspended in deionized formamide and developed on a CEQ 2000XL 8-channel capillary DNA sequencer (Beckman-Coulter).

BINDING ASSAYS

Individual clones amplified from the colony PCR reactions using standard pool primers, in this case 42.40 and 20.40. The PCR reactions are ethanol precipitated, resuspended in water and quantitated on a 4% agarose gel using 100 bp quantitation standards (Gensura). The dsDNA are then transcribed in the presence of α-32P-labeled UTP. The generated RNAs are purified on an 8% polyacrylamide gel, eluted and ethanol precipitated. The concentration of the RNA products is determined using a Spectrophotometer.

Aptamers are assayed for their affinity to the selected target Lysozyme as previously described [18] using a nitrocellulose (Schleicher & Schuell, Keene, NH) and nylon membrane (Amersham Pharmacia Biotech) sandwiched by a Minifold I filtration manifold (Schleicher & Schuell). The binding reactions conducted in triplicates using with equimolar amounts of RNA and Protein incubated in selection buffer at room temperature. They are filtered through the membranes with the selection buffer (20 mM Tris (pH 7.5), 100 mM NaCl, 5 mM $MgCl_2$). Fraction of the RNA that binds to the target is computed using a PhosphorImager SI (Amersham Pharmacia Biotech) and calculated using the following formula:

$$\text{Fraction of RNA bound} = \frac{\text{RNA captured on nitrocellulose}}{\text{RNA on nitrocellulose + RNA on nylon}}$$

SEQUENCE ANALYSIS

With the sequencing data, the random regions for the N40 selected pool are aligned with All nucleotide and amino acid sequences were aligned using a computer program (MacVector.) and visually inspected before analysis. The rationale for removing the constant regions is because they are static and does not contribute in the analysis of primary sequences variation. Moreover, by just analyzing the random region will provide a fair comparison of the isolated clones to the winning single clone for Lysozyme isolated from the N30 pool [11] because of differences in the constant regions. If the primer regions were left in the analysis, the difference genetic distance between the newly isolated clones from N40 will be overestimated because of the additional 60 base pair difference between the two selection pools. Precautions are taken to ensure the integrity of sequence data as previously recommended [19]. The diversity of the isolated sequences are calculated by analyzing their genetic pairwise Hamming distances, defined as the number of base differences between two sequences [12]. The Hamming distance is calculated based on the following formula:

Hamming distance = (1-fraction of shared sites in two aligned sequences) x100

Once the pairwise distance matrix is generated, principal component analysis is carried out using the software PCOORD (available via the web: http://www.hiv.lanl.gov/content/hiv-db/PCOORD/PCOORD.html), which helps visualize the multidimensional variation in the sequences. PCOORD is useful at finding meaningful patterns in sequence data without any prior knowledge about the evolutionary history. It functions in summarizing the variations in the sequences and provides a plot in a limited number of dimensions. It uses a ordination, which is a method that analyses complex data sets to find the best way to describe multi-dimensional dataset [16].

PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) is used to obtain an overview of data, to find groupings, identify outliers and to generate variables that summarize the main sources of variation in the data and can be used in subsequent analyses. The easiest way to understand PCA is as a projection technique analogous to an ordinary window. The three dimensional reality is projected on the two dimensional window. The window is oriented so that as much as possible of the reality can be seen through it. The central idea of PCA is to extract a few, so-called, principal components or scores (**t**) describing as much as possible of the variation present in the data. The principal components are linear combinations of the original variables and are uncorrelated to each other. They can be imagined as mutually orthogonal lines in the multivariate space. As previously shown [20]:

$$X = t_1p'_1 + t_2p'_2 + t_3p'_3 + ...t_Np'_N + E = TP' + E$$

Here, N is the number of principal components and E is the residual matrix. For each component, **t** represents the scores vector and **p'** the loadings vector. The principal components can be determined using the NIPALS algorithm [21] or by Singular Value Decomposition (SVD) [20]. The scores (**t**) show how the objects and experiments relate to each other. The loadings (**p**) reveal which variables are important for explaining the patterns seen in the score plot and can be geometrically interpreted as the angles between the principal component and the original variables. Analysis of the scores and the loadings facilitates the identification and explanation of important groupings in the data.

REFERENCES

1. Brookfield, J., *Fitness*. Curr Biol, 2002. **12**(21): p. R717.

2. Wright, S., *General, Group and Special Size Factors*. Genetics, 1932. **17**(5): p. 603-19.

3. Fischer KH, H.J., *Spin glass*. 1991, Cambridge: Cambridge University Press.

4. Kauffman, S.A., Perelson, AS, *Molecular evolution on rugged landscapes: Santa Fe Institute studies in the sciences of complexity*. 1991.

5. Fontana, W. and P. Schuster, *Continuity in evolution: on the nature of transitions*. Science, 1998. **280**(5368): p. 1451-5.

6. Schuster, P., et al., *From sequences to shapes and back: a case study in RNA secondary structures*. Proc R Soc Lond B Biol Sci, 1994. **255**(1344): p. 279-84.

7. Salehi-Ashtiani, K. and J.W. Szostak, *In vitro evolution suggests multiple origins for the hammerhead ribozyme*. Nature, 2001. **414**(6859): p. 82-4.

8. Schultes, E.A. and D.P. Bartel, *One sequence, two ribozymes: implications for the emergence of new ribozyme folds*. Science, 2000. **289**(5478): p. 448-52.

9. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands*. Nature, 1990. **346**(6287): p. 818-22.

10. Ellington, A.D. and J.W. Szostak, *Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures*. Nature, 1992. **355**(6363): p. 850-2.

11. Cox, J.C. and A.D. Ellington, *Automated selection of anti-protein aptamers*. Bioorg Med Chem, 2001. **9**(10): p. 2525-31.

12. Hamming, R.W., *Coding and Information Theory*. 2nd ed. 1986: Prentice Hall.

13. Kauffman, S.A. and E.D. Weinberger, *The NK model of rugged fitness landscapes and its application to maturation of the immune response*. J Theor Biol, 1989. **141**(2): p. 211-45.

14. Lee, J.F., Cox, J.C., Collett, JR., Ellington, AD, *Exploring Sequence Space Through Automated Aptamer Selections*. Journal of the Association for Laboratory Automation, 2005. **10**(4): p. 213-218.

15. Malmgren, C., et al., *An antisense/target RNA duplex or a strong intramolecular RNA structure 5' of a translation initiation signal blocks ribosome binding: the case of plasmid R1*. Rna, 1996. **2**(10): p. 1022-32.

16. Higgins, D.G., *Sequence ordinations: a multivariate analysis approach to analysing large sequence data sets*. Comput Appl Biosci, 1992. **8**(1): p. 15-22.

17. Clakson, T., D. D. Gussow, and P. T. Jones, *Characterizing recombinant clones directly by 'PCR screening'*, in *PCR 1*, D.R.a.B.D. Hames, Editor. 1991, Oxford University Press: Oxford. p. 210-212.

18. Jhaveri, S.a.E.A., *In vitro selection of RNA aptamers to a protein target by filter immobilization*, in *Current Protocols in Nucleic Acid Chemistry*, e.a. S.L. Beaucage, Editor. 2000, John Wiley and Sons, Inc.: New York. p. 9.3.1-9.3.25.

19. Learn, G.H., Jr., et al., *Maintaining the integrity of human immunodeficiency virus sequence databases*. J Virol, 1996. **70**(8): p. 5720-30.

20. Joliffe, I.T., *Principal Component Analysis*. 1986, New York.

21. Wold, S., *Estimation of principal component and related models by iterative least squares*. Multivariate Analysis, ed. P. Krishnaiah. 1966, New York: Academic Press.

CHAPTER 5: APTAMERS AS DIAGNOSTIC TOOLS FOR LUNG CANCER

INTRODUCTION

LUNG CANCER

Lung cancer is the most prevalent cancer in the world. It is divided into non-small cell lung cancer (NSCLC), which accounts for about 80% of the cases; and small cell lung cancer (SCLC) comprising the other 20% [1]. Most of the lung cancers are grouped under NSCLCs. They are grouped because treatment regimen and prognosis are similar. There are 3 predictors, termed TNM (tumor, node, metastasis) state at diagnosis, that are used for survival prediction of NSCLCs [2]. Available treatments are usually ineffective or excessively toxic. SCLC is considered the more aggressive type of lung cancer with very poor prognosis. Metastases start from the lymph nodes then quickly spreads to the lung, liver, adrenal glands, bone, bone marrow, and brain. Early detection is crucial for this cancer, but the lack of adequate tumor markers makes it a very daunting task. Hence developing diagnostic tools to identify these primary cancers can highly improve prognosis [3].

Finding biomarkers to aid in early detection and improving diagnosis has been a focus of many researchers [4]. The identification of molecular alteration to all tumors in general as well as signals that can subcategorize the cancers can be useful for diagnosis, treatment regime, and therapeutic assessment.[5]. However, because of the sheer number and the level of heterogeneity across the different lung cancers make the process a highly

complex procedure [6, 7]. Cancer biomarkers can be divided into serum biomarkers, tissue biomarkers and sputum biomarkers [8]. NSCLC and SCLC have slightly different biomarker expressions. Neuroendocrine cells such as neuron-specific enolase (NSE), chromogranin A (ChrA), bombesin-like gastrin-releasing peptide, and BB isoenzyme of creatine kinase are all specific to SCLC [9]. For NSCLC, most of the makers include proteins specific for cell cycling, growth factors and their receptors [10, 11].

APTAMER AND DIAGNOSTICS

Aptamers are very effective tools that can be applied in the form of small-molecular detection probes, target inhibitors or target binders. Their relatively small nature allows for easy manipulation and conjugation of these aptamers with other molecules such as quantum dots, oligonucleotides, nanoparticles and siRNA delivery[12, 13]. Aptamers also serve as excellent biomarker sensors due to their highly specific nature [14]. The ability to select aptamers against simple or complex targets and their seemingly pluripotent binding abilities make them excellent reagents for the study, diagnosis, and perhaps even therapy of cancers, as reviewed by Cerchia et al.[15]. In fact, selection experiments have yielded RNA and DNA aptamers that can bind to cell surface targets on tumors [16-19]. In each instance, the aptamers were able to selectively bind transformed cells but not normal cells. For example, Hicke et al., [18] targeted human U251 glioblastoma cells with a 2'-fluoropyrimidine, modified RNA pool. Selected, modified RNA aptamers were found to bind the extracellular matrix protein Tenascin-C (TN-C), a protein that is believed to be a hallmark of the onset and metastasis of cancer. The aptamers formed complexes with TN-C that had a dissociation constant of

5 nM. The aptamers also bound tumor tissue expressing Tenascin-C, but did not bind tissue that lacked TN-C. In fact, the aptamer was also able to discriminate (by 20-fold) against mouse TN-C, even though this protein shared 98% sequence identity with the human protein.

Another example, Lupold et al. [19] targeted prostate-specific membrane antigen (PSMA), which is a membrane-bound glycoprotein that is found in prostate epithelial cells and overexpressed by prostate cancers. In fact, its detection in increased quantities is considered a hallmark of the disease. Two aptamers were selected from a 2'-fluoropyrimidine modified RNA library that spanned 40 random sequence positions. The aptamers were distinct from one another and likely bound different epitopes of the PSMA antigen. The aptamers could inhibit the peptidase activity of PSMA with K_i values in the 2-10 nM range. Fluorescently labeled aptamers were also shown to bind to a prostate tumor line (LNCaP) that normally overexpresses PSMA, but not to PC-3 cells, a different prostate tumor line that lacks this antigen expression [20, 21].

Blank et al. [16] [16] carried out a selection against rat endothelial cells immortalized with adenovirus, and isolated a single-stranded DNA aptamer that proved to be specific for glioblastoma. The aptamer was found to bind to the protein pigpen, which plays a role in angiogenesis by regulating endothelial cell proliferation. Pigpen expression is up-regulated in actively dividing cells and is down-regulated as they become confluent, suggesting that pigpen helps regulate endothelial cell differentiation [22]. The aptamer were able to label the microvessels surrounding the rat glioblastoma, but not microvessels in normal, mature brain vasculature. As was the case with anti-PSMA aptamers, the anti-pigpen aptamers could be fluorescently labeled and used for

contrast staining of transformed cells versus normal cells. Finally, DNA aptamers that were selected against leukemia cells were specific for T-cells and not B-lymphoma cells [23].

Such a diversity of biomarkers exist among lung cancer lines that honing down on a particular target for the disease is very difficult. However the ability of aptamers to target broad range of targets make them great candidates for this purpose. Selecting aptamers against lung tumor lines might seem like a leap of faith because of cell heterogeneity. Aside from such difference, there is also a great degree of similarity among the cells, especially SCLC, NSCLC and normal lung lines. The idea that certain growth factors are overexpressed is no longer a feature to cancerous cells [2]. For clinical applications, the challenge is to develop biosensor that can distinguish among different type of lung cancer, but not so unique that it will only be able to identify a single type of cell line, which will highly limit its utility.

CELL LINE SIGNIFCANCE

Three lung tumor cell lines that differ in their Epidermal Growth Factor Receptor (EGFR) expression were chosen as targets for selection. NCI-H358 is from a non-small cell lung cancer and expresses wild type EGFR. A brief summary of the cell lines is shown in **Table 4.1**. NCI-1650 is from a bronchoalveolar carcinoma and expresses a mutant form of EGFR (in-frame deletion delE746-A750). NCIH526 is from a small cell lung cancer (SCLC) cell lines and expresses low to undetectable levels of EGFR. The first two lines are adherent, but NCI-H526 is a suspension culture. Many agents available target the EGFR and have made more promising results of NSCLC treatment. In addition

to the monoclonal antibody treatments, two available EGFR tyrosine kinase inhibitors, erlotinib and gefitinib, have elicited good response in patients with NSCLC relapse [24, 25]. When combined with chemotherapy, patients under erlotinib had a longer survival [26]. **Figure 5.1** is a summary of the different targets for the therapy that are available up to date for EGFR.

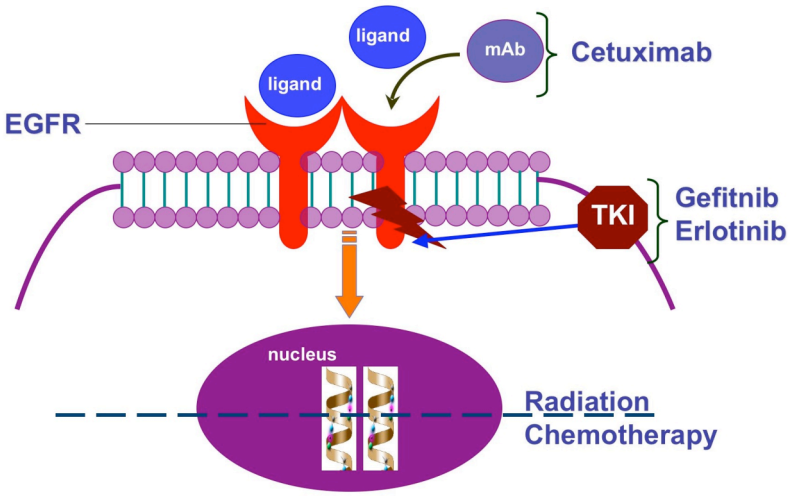


Figure 5.1. Therapeutic Agents Against EGFR. Multi-targeted therapy is one of the golden rules against lung cancer. Different therapeutic agents are utilized in conjunction to target different areas of the cell. Monoclonal antibodies (Cetuximab) compete with extracellular ligands to prevent the dimerization of the EGFR. New intracellular inhibitors for the tyrosine kinase domain (Gefitnib and Erlotinib) stop downstream signaling cascades. Finally, Radiation/Chemotherapy targets the nuclear cellular functions stopping cell cycling activities.

All these facts are the premise for our choice of lines to start our investigation of aptamer selection against lung cancer cells.

	H526	H358	H1650
Organ	Lung	Lung	Lung
Cancer	Carcinoma; variant small cell lung cancer	bronshioalveolar carcinoma; non-small cell lung cancer	Adenocarcinoma; bronchoalveolar carcinoma
Properties	Undetectable levels of EGFR	wild-type EGFR	EGFR (DelE746A750)
Morphology	Suspension	epithelial	epithelial
Origin	bone marrow	bronchiole; alveolus	pleural effusion

Table 5.1. Cell lines Used for Selections

Results and Discussion

The first step in developing biosensors for such a high variety of lung cancer cells is the being able to isolate aptamers that will be specific for each type of lung cancer class. In this report work, we tailor make a selection protocol to select aptamers that can distinguish between cancer cells and normal lung cells; and aptamers that can differentiate between NSCLC and SCLC using a 2'F modified RNA aptamer pool with 30 nucleotides in random region.

Our first approach was conducting direct selection against individual cells. The process of cell surface selection against epithelial cells is illustrated in **Figure 5.2.** A pool of nuclease resistance RNAs are incubated with cells that are grown to about 85% confluence for 3 hours. The cell undergoes a series of wash steps with PBS to rid of the unbound RNA aptamers. The winners are isolated by trypsinizing the cells and washed

further with PBS. The isolated cells are then lysed to isolate any remaining RNAs that may have internalized. The isolated RNAs undergo a series of amplification steps to generate the enriched pool for the next round of selection. For suspension cells, the process is similar except for the "(+) trypsin," which is by-passed.

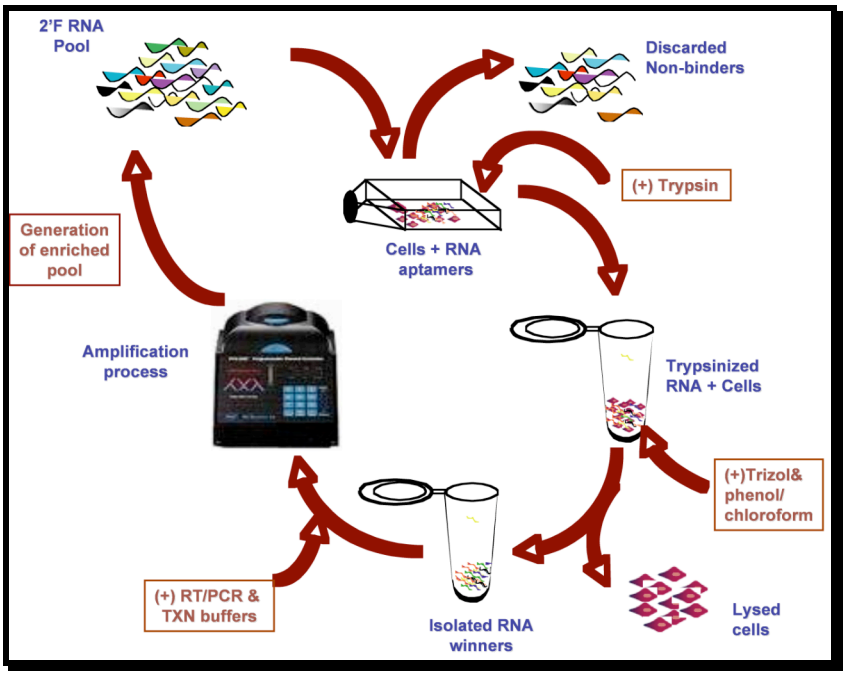


Figure 5.2. Cell Surface Selection Schema.

SCLC (H526) SELECTIONS

Selections against suspension cells H526, which are small cell lung cancer cells, are carried out without for 9 rounds and assayed for binding. The assays are out carried by incubating radiolabeled modified RNAs against the cells. **Figure 5.3** shows the result of H526 selections.

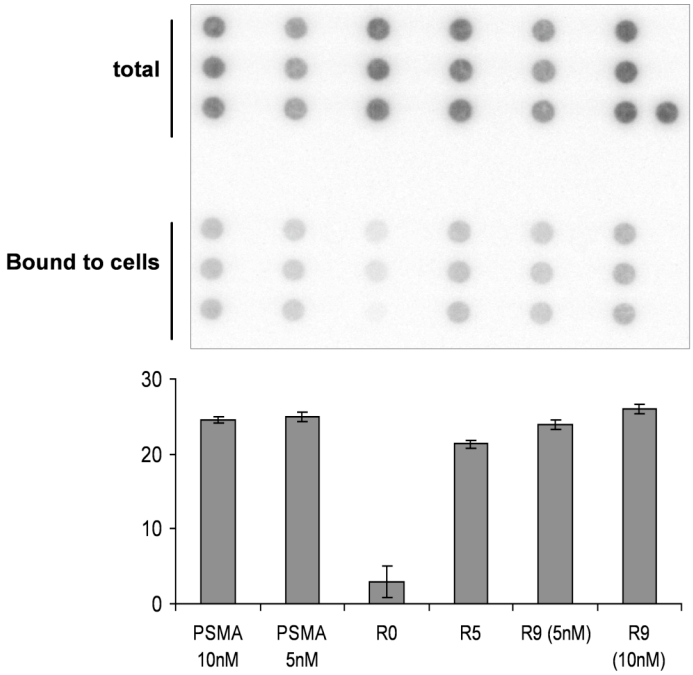


Figure 5.3. Aptamer Selections Against H526. All the clones isolated from round 9 of the selection showed significant affinity towards H526 cells relative to background. Previously selected aptamer A9 [19] was assayed against LnCAP cells to serve as a positive control.

H526 data shows that by round 5, the aptamer pool is showing significant improved affinity towards the cell. Round 9 improvements from round 5 is no longer as drastic. The pool from Round 9 is further sequenced and characterized. The sequences obtained are displayed in **Figure 5.4.** The pool complexity is relatively high. There is no formation of sequence families or common motifs across the different sequences. This diversity in the population is of no surprise because the selection is carried out against a complex target. The first ten clones (clones 1-10) are selected to measure specificity towards H526 and another cell H358, which is a NSCLC cell. The results for this assay are displayed in **Figure 5.5.**

```
>clone-1  GGGAATGGATCCACATCTACGAATTC-CTCGTTTGGAGCGTATTCTACCGTTCACCA-TTCACTGCAGA-TTG
>clone-2  GGGAATGGATCCACATCTACGAATTC-GTTAGTGGTTTTTAGCGACCTCTGCAGCT--TTCACTGCAGA-TTG
>Clone-3  GGGAATGGATCCACATCTACGAATTC-ACCTAATGCAGCATCGTCGCTGTTACAGTA-TTCACTGCAGACTTG
>Clone-4  GGGAATGGATCCACATCTACGAATTC-TAGCTTATACTAGAGCCATTCTTCTCTA---TTCACTGCAGA-TTG
>Clone-5  GGGAATGGATCCACATCTACGAATTC-CATAACATAGCACACTTGGAATTGCTCATG-TTCACTGCAGACTTG
>Clone-6  GGGAATGGATCCACATCTACGAATTC-ACCACAAAGCCGCATTGTCACATGATCCTC-TTCACTGCAGACTTG
>Clone-7  GGGAATGGATCCACATCTACGAATTC-TCACAAGTTTTAGAGTACACTACCCGTC---TTCACTGCAGACTTG
>Clone-9  GGGAATGGATCCACATCTACGAATTC-GCGGCAGCAAGCCGAGTAGGTGCCACTGTG-TTCACTGCAGACTTG
>clone-10 GGGAATGGATCCACATCTACGAATTC-GTGGTATACCTCCCTATGTGCGGCACCTAC-TTCACTGCAGACTTG
>clone-11 GGGAATGGATCCACATCTACGAACTC-CCTGACGCGGCAAGTGTACCGCCTTTCAAG-TTCACTGCAGACTTG
>Clone-12 GGGAATGGATCCACATCTACGAATTC-ACGAAACCCCCACTGCCTATAGCAGTAATG-TTCACTGCAGACTTG
>Clone-13 GGGAATGGATCCACATCTACGAATTC-GCCGGAATATTGACTGAATCGTCAATCTAC-TTCACTGCAGACTTG
>Clone-14 GGGAATGGATCCACATCTACGAATTC-CGGGATTGGATGCCCATGACCCACGGACTA-TTCACTGCAGACTTG
>Clone-15 GGGAATGGATCCACATCTACGAATTC-TAGCTTATACTAGAGCCATTCTTCTCTA---TTCACTGCAGA-TTG
>Clone-16 GGGAATGGATCCACATCTACGAATTC-AAAACGGATGTCGGCATAGATGGAAGCTGG-TTCACTGCAGACTTG
>Clone-17 GGGAATGGATCCACATCTACGAATTC-CGGGTTCTCAGAATACGCTTGACCTACGTG-TTCACTGCAGACTTG
>clone-18 GGGAATGGATCCACATCTACGAATTC-AACTCGCGATAGTGCACGCGCGAATACCCC-TTCACTGCAGACTTG
>clone-19 GGGAATGGATCCACATCTACGAATTC-CCTCAATCAGGCGTAATCCGAACCATTTCG-TTCACTGCAGACTTG
>Clone-20 GGGAATGGATCCACATCTACGAATTC-ACCACTCCACAGAGATCACCTAGAGCCCAC-TTCACTGCAGACTTG
>Clone-21 GGGAATGGATCCACATCTACGAATTC-ACCTAATGCAGCATCGTCGCTGTTCAGTA--TTCACTGCAGACTTG
>Clone-22 GGGAATGGATCTACATCTACGAATTC-TAAGGACAAGTTAGAAATAGGTCAGCGCTG-TTCACTGCAGACTTG
>Clone-23 GGGAATGGACCAAAAGGA-CGAATTC-CTACGTACGGGATCGCTGATACGCGTGTTG-TTCACTGCAGACTTG
>Clone-24 GGGAATGGATCCACATCTACGAATTC-ATAGACGACCGTCTACCCATCGCGGTCAGG-TTCACTGCAGACTTG
>Clone-25 GGGAATGGATCCACATCTACGA--TC-CATAGATTTTAACCGCAGCGCTACATCTTCGTTCACTGCAGACTTG
>Clone-26 GGGAATGGATCCACATCTACGAATTC-AACCCGTACCCTCAGTGCGTACACTACTA--TTCACTGCAGACTTG
>Clone-27 GGGAATGGATCCACATCTACGAATTC-GGAACATGGACCCGCAGCGTTACGTTAATA-TTCACTGCAGACTTG
>Clone-28 GGGAATGGATCCACATCTACGAATTC-CCGTATTCAGCCTAGGATGTAGGCAAGTTG-TTCACTGCAGACTTG
```

Figure 5.4. H526 Selected Clones at Round 9. The primer binding regions are colored maroon and the random regions are in black. The graph shows that the pools remains highly complex and there is no clear appearance of motifs or the formation of specific sequence families.

All clones are assayed in triplicate to generate the error bars. Every clone shows a significant increase in affinity to the targeted cell relative to background. However, none of the clones binding is a prominent as that of clone 2. In addition, the isolated clones are also incubated with H358 to test whether they show any affinity to a NSCLC cancer cell. The data shows no specificity towards the non-selected target at all. It appears that the aptamers selected are specific to the SCLC cells (H526).

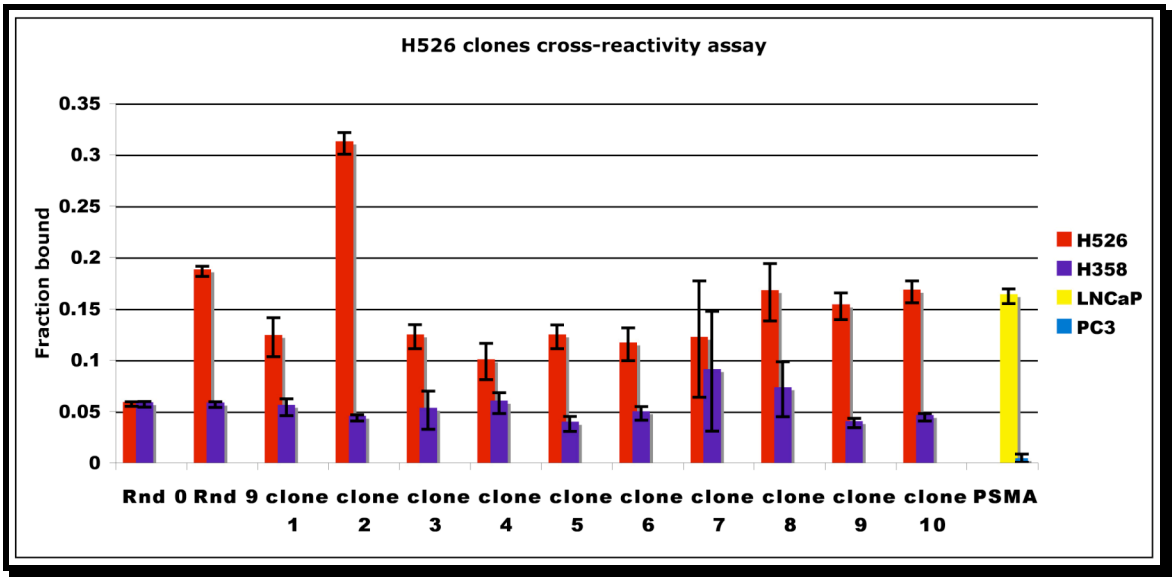


Figure 5.5. Cross Reactivity Assay for H526 Selected Aptamers. The assay shows the binding affinity of individual clones isolated from the round 9 selection. Every clone shows higher binding to the specific cell relative to background. However, clone 2 shows the greatest improvement in binding. Moreover, most of the clones show little to now binding to the non-specific cell H358. Similar to Figure 5.3, PSMA aptamer is used as a positive control for the assays.

Based on these binding assays, the next step is to assay these clones to test their ability to differentiate between NSCLC and SCLC cells. The clones selected for the next set of assay are picked based on their affinity to H526; only the 6 best binders are selected. Additional cell lines are added to further evaluate these clones specificity. Two additional SCLC cells lines along with 4 different NSCLC cell lines are used in the test. **Figure 5.6** shows an overview of the cells used in the assay for determine verify these clones specificity towards SCLC. In summary, each individual point tested contains a cocktail of equal amounts of the different cells.

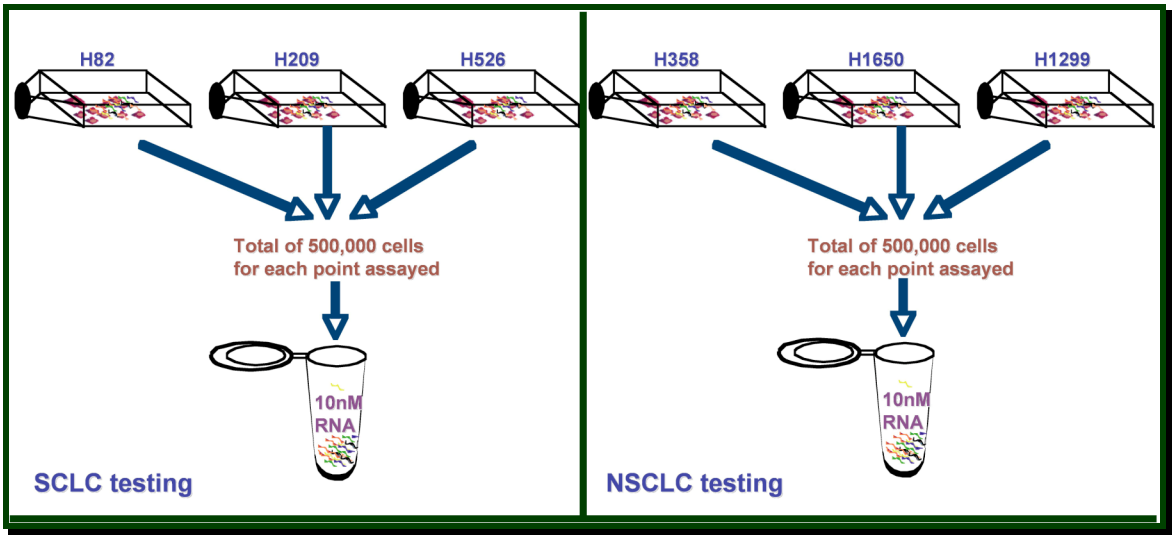


Figure 5.6. Assay Set-up for Aptamers against SCLC and NSCLC. A cocktail composed of 3 different cell lines for each category of lung cancer cell is used for determining the binding affinity of each individual clone.

The results for the individual clone assays from round 9 against a cocktail of SCLC and NSCLC are shown in **Figure 5.6.** All the clones tested here except for clone 8 showed high specificity towards the cocktail of SCLC cell lines versus the NSCLC cocktail. Such a high degree of specificity towards SCLC as opposed to NSCLC can be attributed to the biomarkers that are specific for each type of lung cancer. For example, neuroendocrine cell markers for the SCLC [9] and growth factors for NSCLC to name as previously described [10, 11]. According to the data, it appears as though the great distinctions among these two type of cells is different enough that isolating aptamers that would distinguish between these two cancer categories did not need the introduction of negative selections as it is expected in most of the cell surface selections.

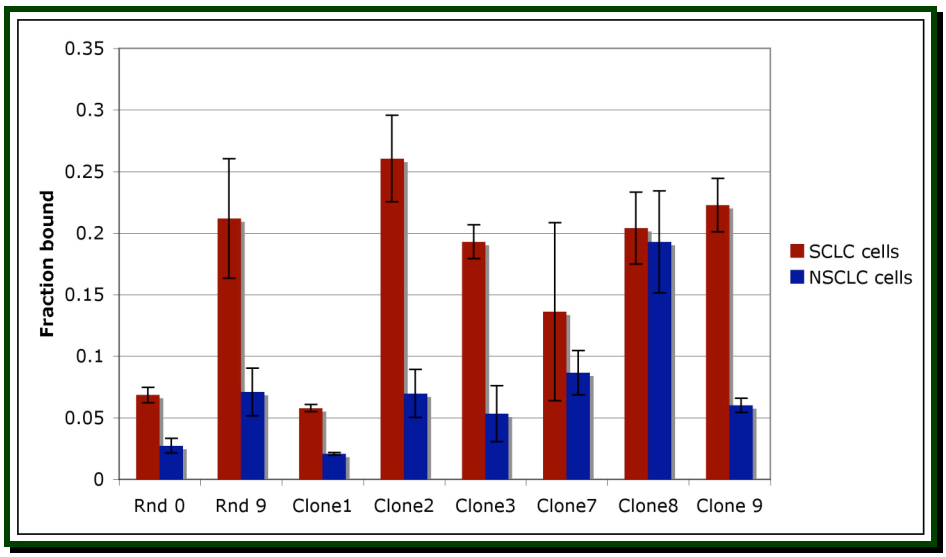


Figure 5.7. Aptamers Against and SCLC and NSCLC. Of the clones tested, only Clone 8 did not distinguish between SCLC cocktail and NSCLC cocktail of cells. Each cocktail contains 3 different cell lines.

SELECTIONS AGAINST H358 AND H1650 (NSCLC) CELLS

Initially selections are carried out as described in **Figure 5.2** for each individual cell. Selections against H358 and H1650 are assayed at after reaching round 15 and 8 respectively. An assay of the last round of selection of each pool is assayed against different NSCLC cell lines and normal lung cells. The data is displayed in **Figure 5.8.** A former graduate student Ted Chu has conducted selections against A549. All pools have binding above that of background against all cells. However, because of the amount of homogeneity among the NSCLC cells and even the normal cells, the pool is unable to differentiate between these cells. Even though, the RNA pool selected against A549 has higher specificity towards the selected target, the pool's affinity towards the other cells remains significantly above background. To further investigate whether we can differentiate between the cells, selections against H358 are used for subsequent investigations.

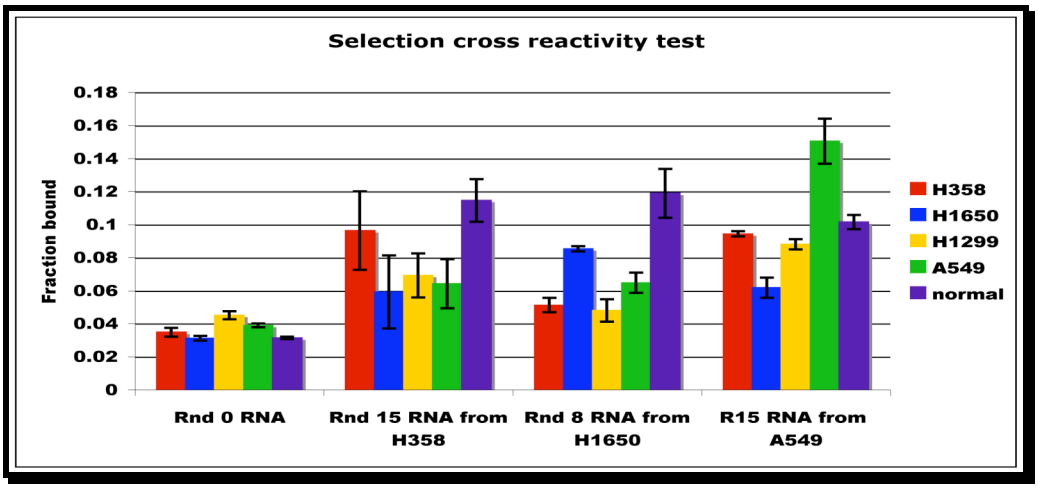


Figure 5.8. Cross Reactivity Testing. The last round of selection for each pool is labeled and assayed against a variety of cells. The figure shows that there is not a significant level of discrimination of the aptamer pool against other cell types that were not selected for. The high level of homogeneity between the variety of NSCLC and normal lung cells makes it impossible for the RNA pools to differentiate them. The "normal cell" is a cocktail of 3 different normal lung lines: MRC-9, NHBE, and CCD-Lu16.

More tailoring of the selection was further conducted to isolate aptamers that could specifically differentiate between normal and cancer cells. Understanding how to customize selections for a specific purpose is crucial in developing aptamers especially for such a complex target that can both be homogenous and heterogeneous. Starting back from round 10 of the RNA that have been selected against H358, additional selections against H358 were conducted. Negative selections against another NSCLC cell (H1650) and normal lung cell lines were incorporated into the selection as illustrated in **Figure 5.9.**

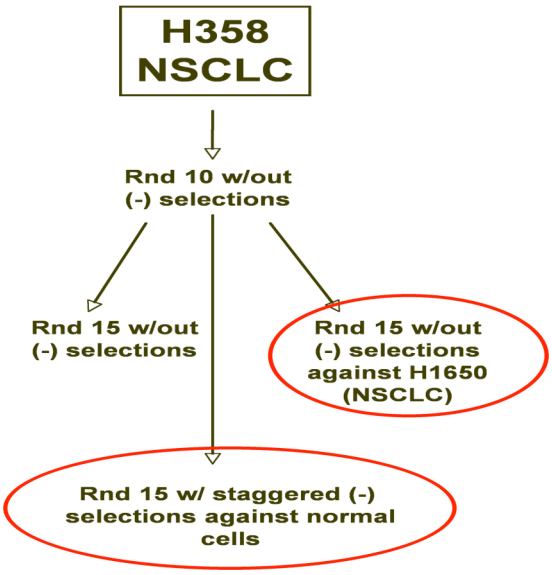


Figure 5.9. Additional H358 Selections. Additional selections are introduced to tailor generate aptamers that can distinguish particular lines of interest (circled in red). Instead of restarting the selections from round 0, all new selections are initiated from round 10 of the selected pool. The population of pool in round 10 is still very diverse as verified through sequencing of the pool.

The purpose of introducing negative selections was to investigate the degree of selection fine-tuning that was needed to isolate the aptamers. This was important because of the large variation and similarity among lung cancer and normal cells. Negative cell selections involved an extra step as shown in **Figure 5.10.** The initial incubation of the pool took place with the non-specific target; in this case it would be H1650 or a cocktail of normal lung cell lines. The cells were then washed and the non-binders were then incubated with H358, the targeted cells. The selection then proceeded as described earlier.

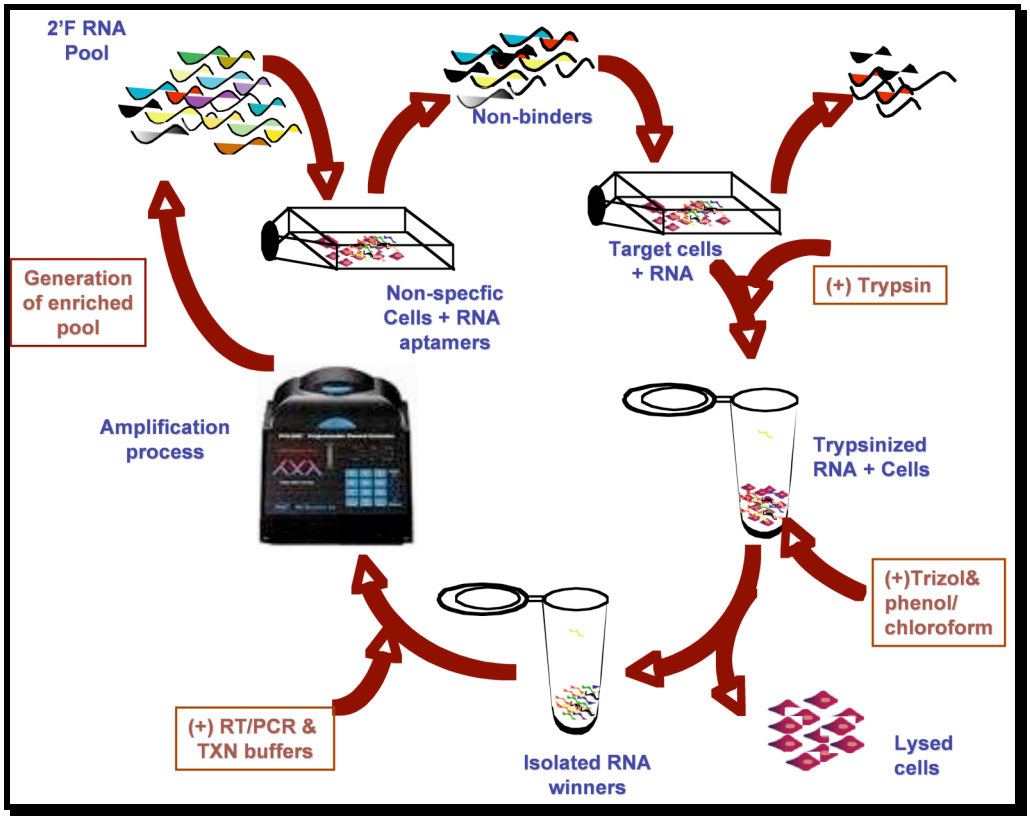


Figure 5.10. Negative Selection of Aptamers Against Cell Surfaces. Negative selections involve the incubation of the pool with the non-targeted cells, and the non-binders recuperated from there are then incubated with the target of interest.

The first negative selection involved selecting for aptamers that could distinguish between two different NSCLC cell lines as a proof of principle for the efficiency of isolating aptamers with particular specificity. The results are summarized in **Figure 5.11.** As the illustration shows, at round 10 of the selections against H358 and before the introduction of negative selections, the pool has an increased affinity towards H358 cells relative to round 0. However, this pool also displayed increased binding towards non-specific cells such as H1650, H526, and normal lung cell lines. Another interesting observation was that selections against this NSCLC cell yielded a pool that had an

affinity towards the SCLC cell, H526. At round 15, after the introduction of negative selections against H1650, the affinity of the final pool towards H1650 was highly reduced as expected. Moreover, the pool's affinity towards the SCLC cells was reduced as well. A plausible explanation for the sudden reduction in binding towards H526 could be attributed to the event that the negative selection reduced the population complexity and the amount of the pool going into the targeted cell. This event allowed for the increase in the stringency of the selection indirectly against H526 cells. The obvious differences between NSCLC and SCLC were prominent enough that by merely increasing the selection pressure of reducing the pool size was sufficient to isolate a pool that preferred the one cell over the other without negative selections. Unfortunately, the round 15 pool was unable to distinguish between H358 and normal cells. It could be inferred that normal cells shared more similarities with NSCLC cells than with SCLC cells. It has been shown previously that no morphological differences were observed between NSCLC lung mouse lung cell lines and normal mouse lung lines [27].

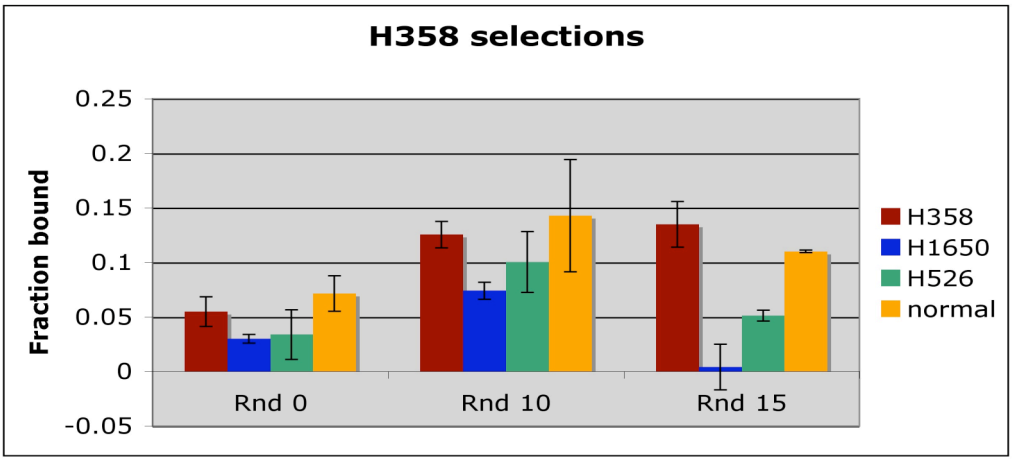


Figure 5.11. H358 Selections with Negative Selections against H1650 Cells. The introduction of negative selection against H1650 cells highly reduced the round 15 pools

ability to recognize H1650 cells. Normal cells are composed of a cocktail of MRC-9, CCD-Lu16, and NHBE cells.

The next step was to do H358 selections with the introduction of negative selections against a cocktail of normal cells. Being able to isolate aptamers that can distinguish between cancerous versus non-cancerous cells has more clinical relevance than purely distinguishing between different cell lines. The results from this selection are shown in **Figure 5.12.** The assays showed that before the introduction of negative selections against normal lung cell lines at round 10, the pool was unable to differentiate between these cell lines. However, after negative selections were introduced, the pool showed increased specificity towards the targeted cells H358.

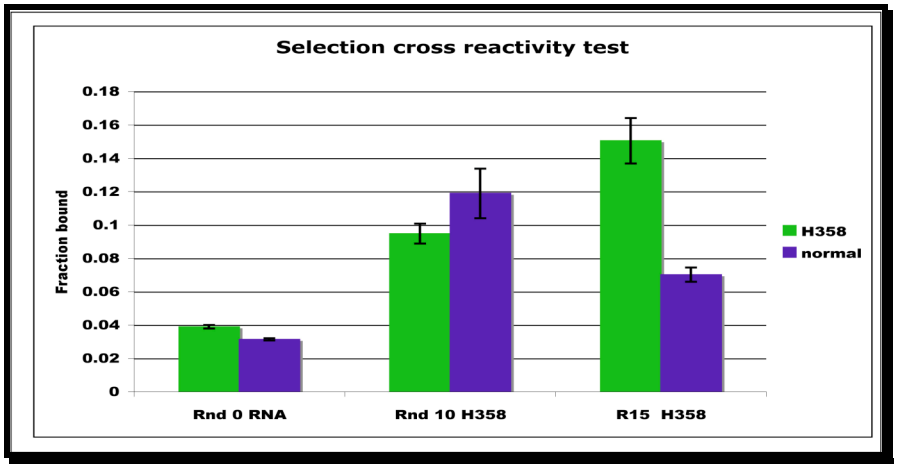


Figure 5.12. Binding Assays of H358 Selections. The negative selections against the normal cells were introduced after round 10. At round 15 the pool was labeled and assayed for specificity against H358 and the cocktail (MRC-9, CCD-Lu15, NHBE) cocktail of normal cells, the pool showed significantly higher specificity towards H358 cells.

The round 15 pool of this selection was cloned and sequenced. The sequences obtained are shown in **Figure 5.12.** Similar to H526 selections, the isolated clones were all distinct form each other and no common motifs were found. This was also an indication

of the level of heterogeneity and complexity of the targets. The first 10 individual clones (highlighted in yellow in **Figure 5.13)** were further tested for specificity. The results are shown in **Figure 5.14.** With the exception of two clones isolated, E7 and D6, the majority of the clones showed a high degree of preferential binding to H358 cells. The purpose of isolating these clones was to potentially serve as biosensors for the detection of tumor cells versus normal cells for our collaborators at MD Anderson. The next step that they are planning on undertaking is to label these aptamers and test them against their home made tissue arrays.

```
> E7 GGGAATGGATCCACATCTACGAATTC-GCGATCGTAGGTGCACACGCAAGACATGGG--TTCACTGCAGACTTGACGAAGCTTA
> D7 GGGAATGGATCCACATCTACGAATTC-GTAGGTTTCGGTCGTGGCAACCGTACCAGC--TTCACTGCAGACTTGACGAAGCTTA
> G7 GGGAATGGATCCACATCTACGAATTC-GTACACAAGCTGTTTGGCACAGACCTTCCG--TTCACTGCAGACTTGACGAAGCTTA
> E5 GGGAATGGATCCACATCTACGAATTC-GTATGCACGTCAGCCAACGCTGCGTTTGAC--TTCACTGCAGACTTGACGAAGCTTA
> G6 GGGAATGGATCCACATCTACGAATTC-TGCACGTAGGTGACGGATTGCCCACGCGAC--TTCACTGCAGACTTGACGAAGCTTA
> A6 GGGAATGGATCCACATCTACGAATTC-ACGGATTGTGGAGTTGATTGCGGCGTCATG--TTCACTGCAGACTTGACGAAGCTTA
> B6 GGGAATGGATCCACATCTACGAATTC-TGGATGACCACTGACCCACAATCGCTCATG--TTCACTGCAGACTTGACGAAGCTTA
> D5 GGGAATGGATCCACATCTACGAATTC-CCAGTGGCTAGTCCCCCTTGTGCAGCACCA--TTCACTGCAGACTTGACGAAGCTTA
> H5 GGGAATGGATCCACATCTACGAATTC-CGATGGACTGTCCCGATGCAGTTTGGTGTG--TTCACTGCAGACTTGACGAAGCTTA
> F5 GGGAATGGATCCACATCTACGAATTC-GATACCATTCAACAAGTTTGTATAGCGCCG--TTCACTGCAGACTTGACGAAGCTTA
> C5 GGGAATGGATCCACATCTACGAATTC-ACAGTTTCCTTACAGTGCGACTTTCTCGAT--TT-ACTGCAGACTTGACGAAGCTTA
> A7 GGGAATGGATCCACATCTACGAATTC-ACGTCCAAGTTTCCGCAACGTAGTGTAGTG--TTCACTGCAGACTTGACGAAGCTTA
> F6 GGGAATGGATCCACATCTACGAATTC-ACCG-CGTAGATCATCCTAGCCGGGCATGTA-TTCACTGCAGACTTGACGAAGCTTA
> H6 GGGAATGGATCCACATCTACGAATTC-GGTAACGTAGATGGCATTGCATTGTGGTGG--TTCACTGCAGACTTGACGAAGCTTA
> C7 GGGAATGGATCCACATCTACGAATTC-GAACTAACTTAGTGGTTAGTCGCTGCACTG--TTCACTGCAGACTTGACGAAGCTTA
> C6 GGGAATGGATCCACATCTACGAATTC-AAGGTGTGGCCTGGCAGCGATATTGTGGTA--TTCACTGCAGACTTGACGAAGCTTA
> G5 GGGAATGGATCCACATCTACGAATTC-ACCGATGCCACGTACCTTTTGTACGCTCCA--TTCACTGCAGACTTGACGAAGCTTA
> B5 GGGAATGGATCCACATCTACGAATTC-CTCGGGTTCACAGCTGAGCACCAGCCCATG--TTCACTGCAGACTTGACGAAGCTTA
```

Figure 5.13. Individual Clones from H358 Selections with Normal Cells as Non-specific Targets. Constant regions are in maroon and the N30 random region are colored black. No motif families are found among the sequences isolated from this selection.

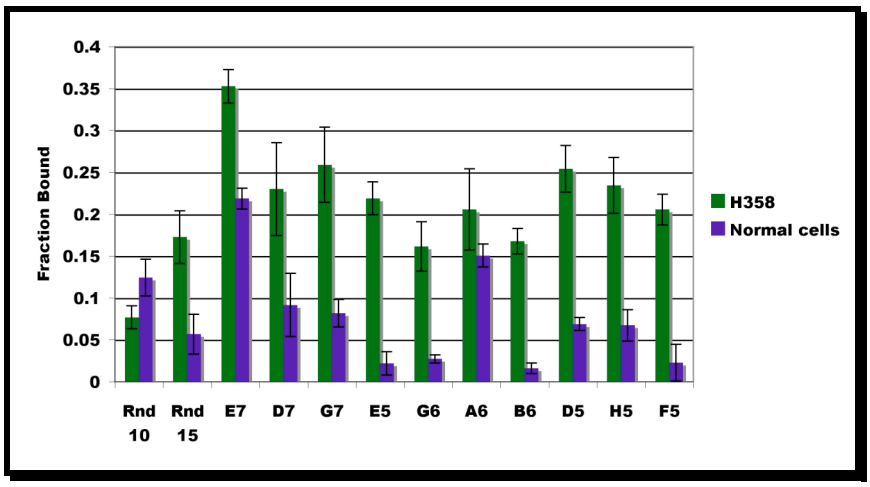


Figure 5.14. Round 15 Binding Assay Profiles. Individual clones from the R15 selection of H358 cell with negative selections against normal cells show high specificity towards the selected cells with the exception of two clones.

We have been able to isolate aptamers that were able to distinguish between SCLC and NSCLC cells by doing a traditional cell surface selection against an SCLC cell (H526). These aptamer can potentially function as biosensors for the early detection of cancer cells and identify whether the cancerous cells are SCLC or NSCLC. The importance of this type of distinction is that it will allow clinicians to apply appropriate therapy. Because of the uniqueness biomarkers of SCLC cells, no negative selection was needed to isolate the aptamers that could discriminate against other cells. When tested against other SCLC cells, some of the aptamer were also able to recognize them. This makes them very powerful tools, because being able to identity a particular category of cancer is useful for clinical aspects than being able to identify one single cell line. However, more work should be involved the evaluation of a larger number of SCLC cell lines to determine whether the aptamers ability to recognize these cells are global. Testing these aptamer on human lung tissues will further support their potency for

clinical use. Due to the limitations in the lab, our collaborators from MD Anderson will be conducting these studies.

From the NSCLC cell selections, the outcomes were not as clean cut. Since NSCLC makes up more than two thirds of the lung cancers, the level of heterogeneity among them is obvious. The aptamers isolated from positive selections alone had a high level of cross-reactivity towards other cancerous and non-cancerous cells. Such wide range of recognition renders these aptamers useless. We tested how negative selections could alter the outcomes of the aptamers isolated. From our work, the introduction of a non-specific targeting step produced aptamers that were able to highly discriminate between these cells but not others. For example, the negative selections conducted against H1650 yielded aptamers that were highly specific to H358 and not H1650. However, these aptamers also recognized normal lung cell lines. By introducing negative selections against normal cell lines, we were able to isolate specific aptamers that recognized H358 with high specificity. These aptamers were generated with the purpose of having them tested against tissue arrays that have been generated by Dr. Wistuba at MD Anderson to assess whether they will be able to distinguish between normal and NSCLC cell in tissue.

CONCLUSIONS

We have been able to tailor the selection process to isolate specific aptamers for our own purpose. However, there are still a lot of cells and more aptamers from specific cancer types may be needed to serve in lung cancer diagnostics. Another challenge is that most aptamers that are carried out against cell lines might not be functional against

isolated tissues. The testing for the isolated aptamer from these selections against tissue samples will further validate the method that we have improved.

The versatility of tailoring aptamers to specific targets is well known. Using aptamers to target cell surface markers have been conducted in various manners included direct cell surface selections [28-30], or selecting a purified cell surface protein [16, 19, 31]. There are advantages and disadvantages to both forms of aptamer selections. The downside of directly selecting against cell surface is that the aptamers' targets are not known, and additional steps are needed to determine the target. Selecting against purified protein may produce aptamers specific to the protein; however, they might not recognize the protein in its native form found on cancer cell surfaces.

There are a large number of variables that can affect the success of a selection experiment. Very few selections target heterogeneous whole cells surfaces and thus the parameters that influence the selection are not well understood. Understanding and controlling these variables will help maximize the chances of success with a wider range of tumor antigens and cells. The level of complexity and the number of available lung cancer cell lines poses an even greater challenge to isolate specific targets without a detail understanding of the fine-tuning that are required.

FUTURE DIRECTIONS

Because of the large number of selections that are needed to reach an optimal selection method for isolating useful aptamers for lung cancer cells, automating this process will prove to be highly useful. I have developed a protocol that will permit the Tecan Genesis (**Figure 5.15A)** and the Biomek 2000 to carry cell surface of SCLC

suspension cell lines. Letha Sooter had previously configured the Tecan Genesis workstation to carry out ssDNA selections [32]. For the purpose of cell selections a lot of modifications are needed to accommodate the selections. I designed a flow chart of the steps and processes that will take place when cell surface selections are translated into an automated workstation (**Figure 5.15B).** The selection will start with the incubation of the RNA with the cells that are acquired from a source that has refrigeration. Since selections are require about 16 hours to complete 6 rounds of selections, it is not possible to have the cells remain at room temperature for such an extensive period of time. The reaction will then be transferred to a partition area where the non-binders will be removed from the winners in the selection. The winners will undergo a series of amplification steps and a fraction of the RNA isolated will undergo the next round of selection.

Figure 5.15A. Tecan Genesis Workstation. This workstation is outfitted with many third party instruments that are essential to carry out *in vitro* selections. The advantage of

the Tecan over the Biomek 2000 is the ability to carry out selections against 96 targets as opposed to only 8 in parallel.

When designing the steps involved in robotic selections, it is very important to take into account the instruments and consumables that can be outfitted on the worksurface. Moreover, the material of the consumables plays a crucial role for the selection outcomes. Once the steps and the consumables are determined, it is necessary to equip the worksurface with the required components as shown in **Figure 5.15C.** A dry run of this protocol has been conducted already. We currently have a high school student Daniel Winkler, who will continue working on programming the platform and further enhance the process from the existing skeleton found on the robot.

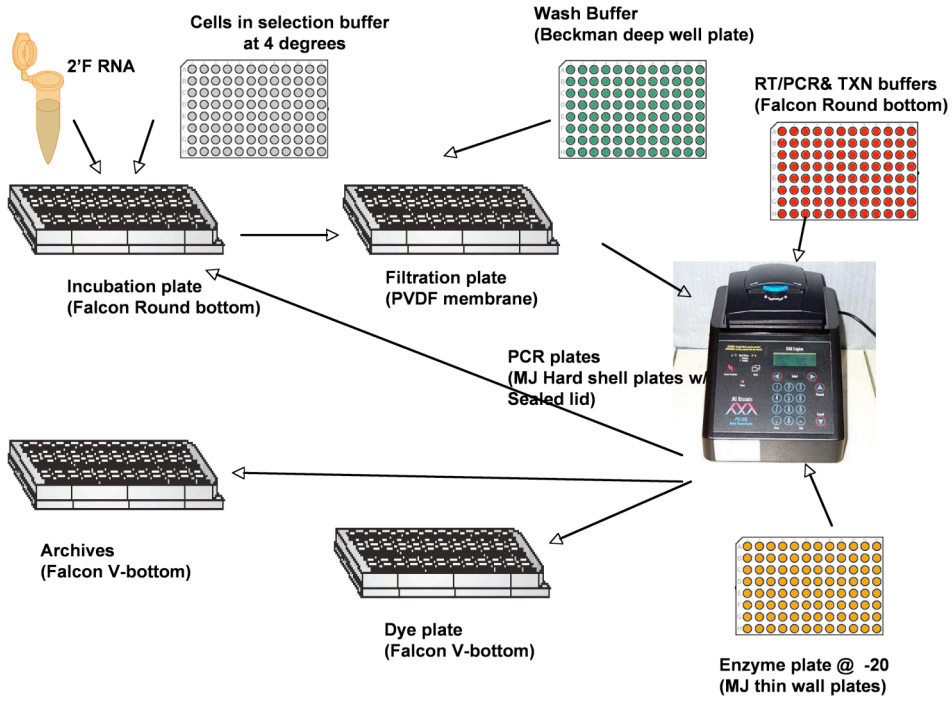


Figure 5.15B. Flow Chart of the Automated Cell Surface Selection Process. The process involves like normal selections the incubation of the RNAs with the target cells.

However, it is imperative to have the cells stored in a place that ensures their integrity for a period of over 16 hours for 6 round of selections. Alternatively, 3 rounds of selections can be carried in half the amount of that time so that new cells can be fed preserving the health of the mammalian cells.

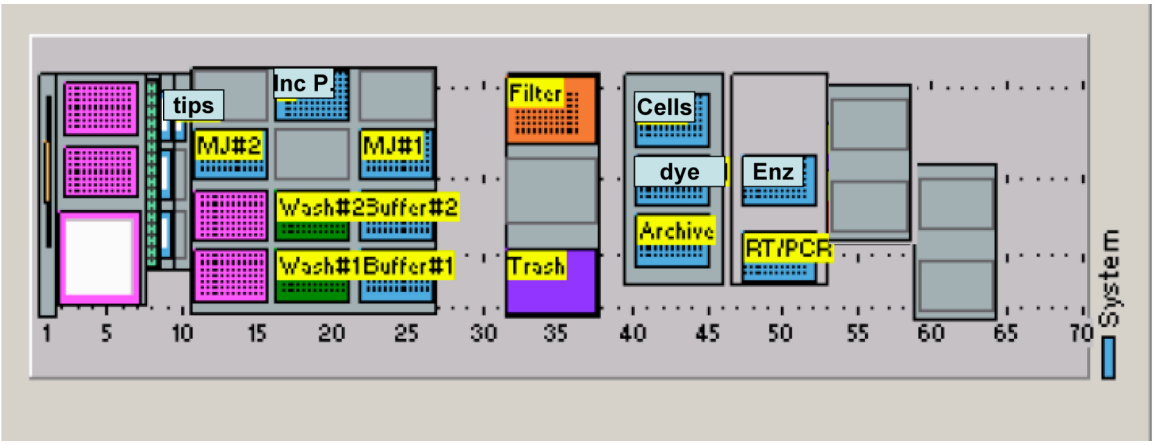


Figure 5.15C. Tecan Genesis Worksurface Layout. Model of where the different plate from Figure 5.15B will be placed on the worksurface. The rationale and critical aspect of the plate locations is to ensure that the movement of the robot arms is as uni-directional (left to right) as possible without the back and forth toggling. This will ensure that less contamination will take place during the selection.

MATERIALS AND METHODS

CELL LINES:

Lung cancer cell lines H358, H526, H1299, H1650, A549, H82, H209 are obtained from ATCC (Rockville, MD). Normal lung lines NHBE, CCD-Lu16, MRC-9 were obtained from our collaborators Dr. Jack Roth at MD Anderson Cancer Research Center. All cells well cultured according to ATCC instructions, except for NHBE, which is cultured according to Clonetics instructions.

IN VITRO TRANSCRIPTION OF APTAMERS AND PURIFICATION:

Aptamer library of DNA with random insert of 30 nucleotides was synthesized following: N30 pool (GGG AAT GGA TCC ACA TCT ACG AAT TCN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNT TCA CTG CAG ACT TGA CGA AGC TT), 41.30 (GAT AAT ACG ACT CAC TAT AGG GAA TGG ATC CAC ATC TAC GA), 24.30 (AAG CTT CGT CAA GTC TGC AGTGAA). The 2'F-RNA library was generated by T7 in vitro transcription in transcription buffer (4% (w/v) polyethylene glycol 8000, 40 mM Tris-HCL (pH 8.0), 12 mM $MgCl_2$, 5 mM DTT, 1 mM spermidine chloride, and 0.002% Triton X-100) with 4 mM 2'Fluorine-CTP (2'F-CTP), 4 mM 2'Fluorine-UTP (2'F-UTP), 1 mM ATP and 1 mM GTP using a Y639F T7 RNA polymerase. Adding DNase then digests DNA templates. Labeled RNA is then purified with an 8% polyacrylamide gel, eluted in water at 37°C overnight, ethanol precipitated then resuspended in water and quantitated with a spetrophometer. For radiolabeled aptamers, 5 μCi of α-^{32}P-labeled GTPs are be added to the transcription.

APTAMER SELECTION:

In the initial cell surface selection, 2'F modified RNA aptamer pool (3 mg) is incubated with cells. After incubation, unbound cells are washed away and the RNAs are isolated by first trypsinizing the cells and then resuspending them in PBS. The cells are then washed 3 times with 1 ml of binding buffer. The cells are then lysed with 300 ml of trizol (Invitrogen, Carlsbad, California) at room temperature for 5 minutes, followed by standard phenol/chloroform extraction, and ethanol precipitation. The recovered aptamer are suspended in reverse transcription/PCR buffer (5X RT/PCR buffer: 10 mM of Tris pH 8.4, 500mM of KCl, 160mM of $MgCl_2$, 5% of Acetamide, 0.5% of Non-idet P40 and

2mM of dNTP). The reaction is denatured at 70^{o}C for 5 minutes and cooled to room. The next step involves the addition of AMV Reverse Transcriptase (USB Corporation, Cleveland, Ohio) and Taq DNA polymerase (Invitrogen, Carlsbad, California). The reaction undergoes a reverse transcription at 50^{o}C for 10 minutes and then PCR amplified for 20 cycles (15 sec at 94 ^{o}C, 15 sec at 45C and 72 ^{o}C at 60 sec). The products are then ethanol precipitated, suspended in water and measured according to efficient coefficient (720700 L/(mole*cm)) at OD260). Transcriptions are further carried out to generate the enriched pool for the next round of selection.

CHARACTERIZATION OF INDIVIDUAL CLONES

DNA generated from selection were cloned, sequenced as previously described [33]. $3x10^{5}$/50 ul of cells in PBS with 5mM $MgCl_2$ are incubated with ^{32}P-labeled aptamers for 30 minutes at room temperature, followed by wash step with 100 uL of PBS with 5mM $MgCl_2$. Cells are then resuspended in 100 uL of PBS with 5mM $MgCl_2$ and filtered through the filtration manifold (Minifold I filtration manifold, Schleicher & Schuell, Keene, NH) with Hybond-N+ nylon transfer membrane (Amersham Pharmacia Biotech). The binding of each aptamer is computed by using a PhosphorImager SI (Amersham Pharmacia Biotech) by comparing the radiation intensity from its original concentration with the signal that remains from cell suspension.

The work presented in this Chapter is not complete due to submission dates of dissertations of the Graduate School. Regardless, the continuation of this work will proceed in a publication-oriented manner. Mass Spec experiments will be carried out to

determine the surface markers that the aptamers are recognizing. Another graduate in the laboratory will continue carrying cell surface selections against lung cancer cells and leukemia cells.

REFERENCES

1. **Brambilla, E., et al., *The new World Health Organization classification of lung tumours.* Eur Respir J, 2001. 18(6): p. 1059-68.**

2. **Bharti, A., P.C. Ma, and R. Salgia, *Biomarker discovery in lung cancer-promises and challenges of clinical proteomics.* Mass Spectrom Rev, 2007.**

3. **Chute, J.P., et al., *Twenty years of phase III trials for patients with extensive-stage small-cell lung cancer: perceptible progress.* J Clin Oncol, 1999. 17(6): p. 1794-801.**

4. **Gail, M.H., et al., *Multiple markers for lung cancer diagnosis: validation of models for localized lung cancer.* J Natl Cancer Inst, 1988. 80(2): p. 97-101.**

5. **Sander, C., *Genomic medicine and the future of health care.* Science, 2000. 287(5460): p. 1977-8.**

6. **Brundage, M.D., D. Davies, and W.J. Mackillop, *Prognostic factors in non-small cell lung cancer: a decade of progress.* Chest, 2002. 122(3): p. 1037-57.**

7. **McCarthy, N.J. and S.M. Swain, *Tumor markers: should we or shouldn't we?* Cancer J, 2001. 7(3): p. 175-7.**

8. **Strauss, G.M. and A.T. Skarin, *Use of tumor markers in lung cancer.* Hematol Oncol Clin North Am, 1994. 8(3): p. 507-32.**

9. **Ma, P.C., et al., *Circulating tumor cells and serum tumor biomarkers in small cell lung cancer.* Anticancer Res, 2003. 23(1A): p. 49-62.**

10. **Shapiro, G.I., et al., *Reciprocal Rb inactivation and p16INK4 expression in primary lung cancers and cell lines.* Cancer Res, 1995. 55(3): p. 505-9.**

11. **Zereu, M., J.J. Vinholes, and C.G. Zettler, *p53 and Bcl-2 protein expression and its relationship with prognosis in small-cell lung cancer.* Clin Lung Cancer, 2003. 4(5): p. 298-302.**

12. **Dwarakanath, S., et al., *Quantum dot-antibody and aptamer conjugates shift fluorescence upon binding bacteria.* Biochem Biophys Res Commun, 2004. 325(3): p. 739-43.**

13. **Farokhzad, O.C., J.M. Karp, and R. Langer, *Nanoparticle-aptamer bioconjugates for cancer targeting.* Expert Opin Drug Deliv, 2006. 3(3): p. 311-24.**

14. **Famulok, M. and G. Mayer, *Chemical biology: aptamers in nanoland.* Nature, 2006. 439(7077): p. 666-9.**

15. **Cerchia, L., et al., *Nucleic acid aptamers in cancer medicine.* FEBS Lett, 2002. 528(1-3): p. 12-6.**

16. **Blank, M., et al., *Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels. selective targeting of endothelial regulatory protein pigpen.* J Biol Chem, 2001. 276(19): p. 16464-8.**

17. **Daniels, D.A., et al., *A tenascin-C aptamer identified by tumor cell SELEX: systematic evolution of ligands by exponential enrichment.* Proc Natl Acad Sci U S A, 2003. 100(26): p. 15416-21.**

18. **Hicke, B.J., et al., *Tenascin-C aptamers are generated using tumor cells and purified protein.* J Biol Chem, 2001. 276(52): p. 48644-54.**

19. **Lupold, S.E., et al., *Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen.* Cancer Res, 2002. 62(14): p. 4029-33.**

20. **Chu, T.C., et al., *Aptamer:toxin conjugates that specifically target prostate tumor cells.* Cancer Res, 2006. 66(12): p. 5989-92.**

21. **Chu, T.C., et al.,** ***Labeling tumor cells with fluorescent nanocrystal-aptamer bioconjugates.*** **Biosens Bioelectron, 2006. 21(10): p. 1859-66.**

22. **Alliegro, M.C. and M.A. Alliegro,** ***Nuclear injection of anti-pigpen antibodies inhibits endothelial cell division.*** **J Biol Chem, 2002. 277(21): p. 19037-41.**

23. **Shangguan, D., et al.,** ***Aptamers evolved from live cells as effective molecular probes for cancer study.*** **Proc Natl Acad Sci U S A, 2006. 103(32): p. 11838-43.**

24. **Fukuoka, M., et al.,** ***Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected].*** **J Clin Oncol, 2003. 21(12): p. 2237-46.**

25. **Kris, M.G., et al.,** ***Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial.*** **Jama, 2003. 290(16): p. 2149-58.**

26. **Shepherd, F.A., et al.,** ***Erlotinib in previously treated non-small-cell lung cancer.*** **N Engl J Med, 2005. 353(2): p. 123-32.**

27. **Malkinson, A.M., et al.,** ***Mouse lung epithelial cell lines--tools for the study of differentiation and the neoplastic phenotype.*** **Toxicology, 1997. 123(1-2): p. 53-100.**

28. **Homann, M. and H.U. Goringer,** ***Uptake and intracellular transport of RNA aptamers in African trypanosomes suggest therapeutic "piggy-back" approach.*** **Bioorg Med Chem, 2001. 9(10): p. 2571-80.**

29. **Lorger, M., et al.,** ***Targeting the variable surface of African trypanosomes with variant surface glycoprotein-specific, serum-stable RNA aptamers.*** **Eukaryot Cell, 2003. 2(1): p. 84-94.**

30. **Morris, K.N., et al.,** ***High affinity ligands from in vitro selection: complex targets.*** **Proc Natl Acad Sci U S A, 1998. 95(6): p. 2902-7.**

31. **Hicke, L., *A new ticket for entry into budding vesicles-ubiquitin.* Cell, 2001. 106(5): p. 527-30.**

32. **Sooter, L.J., et al., *Toward automated nucleic acid enzyme selection.* Biol Chem, 2001. 382(9): p. 1327-34.**

33. **Cox, J.C. and A.D. Ellington, *Automated selection of anti-protein aptamers.* Bioorg Med Chem, 2001. 9(10): p. 2525-31.**

BIBLIOGRAPHY

1. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands*. Nature, 1990. **346**(6287): p. 818-22.

2. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase*. Science, 1990. **249**(4968): p. 505-10.

3. Blank, M. and M. Blind, *Aptamers as tools for target validation*. Curr Opin Chem Biol, 2005. **9**(4): p. 336-42.

4. Proske, D., et al., *Aptamers-basic research, drug development, and clinical applications*. Appl Microbiol Biotechnol, 2005. **69**(4): p. 367-74.

5. Nimjee, S.M., C.P. Rusconi, and B.A. Sullenger, *Aptamers: an emerging class of therapeutics*. Annu Rev Med, 2005. **56**: p. 555-83.

6. Yan, A.C., et al., *Aptamers: prospects in therapeutics and biomedicine*. Front Biosci, 2005. **10**: p. 1802-27.

7. DeAnda, A., Jr., et al., *Pilot study of the efficacy of a thrombin inhibitor for use during cardiopulmonary bypass*. Ann Thorac Surg, 1994. **58**(2): p. 344-50.

8. Bless, N.M., et al., *Protective effects of an aptamer inhibitor of neutrophil elastase in lung inflammatory injury*. Curr Biol, 1997. **7**(11): p. 877-80.

9. Bock, L.C., et al., *Selection of single-stranded DNA molecules that bind and inhibit human thrombin*. Nature, 1992. **355**(6360): p. 564-6.

10. Lehman, N., et al., *The genotypic landscape during in vitro evolution of a catalytic RNA: implications for phenotypic buffering*. J Mol Evol, 2000. **50**(5): p. 481-90.

11. Klug, S.J. and M. Famulok, *All you wanted to know about SELEX*. Mol Biol Rep, 1994. **20**(2): p. 97-107.

12. Burgstaller, P., M. Kochoyan, and M. Famulok, *Structural probing and damage selection of citrulline- and arginine-specific RNA aptamers identify base positions required for binding*. Nucleic Acids Res, 1995. **23**(23): p. 4769-76.

13. Mannironi, C., et al., *Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif*. Rna, 2000. **6**(4): p. 520-7.

14. Fontana, W. and P. Schuster, *Shaping space: the possible and the attainable in RNA genotype-phenotype mapping*. J Theor Biol, 1998. **194**(4): p. 491-515.

15. Fontana, W., et al., *RNA folding and combinatory landscapes*. Physical Review. E. Statistical Physics, Plasmas, Fluids, And Related Interdisciplinary Topics, 1993. **47**(3): p. 2083-2099.

16. Schuster, P., et al., *From sequences to shapes and back: a case study in RNA secondary structures*. Proc Biol Sci, 1994. **255**(1344): p. 279-84.

17. Schultes, E.A. and D.P. Bartel, *One sequence, two ribozymes: implications for the emergence of new ribozyme folds*. Science, 2000. **289**(5478): p. 448-52.

18. Held, D.M., et al., *Evolutionary landscapes for the acquisition of new ligand recognition by RNA aptamers*. J Mol Evol, 2003. **57**(3): p. 299-308.

19. Cox, J.C. and A.D. Ellington, *Automated selection of anti-protein aptamers*. Bioorg Med Chem, 2001. **9**(10): p. 2525-31.

20. Sooter, L.J., et al., *Toward automated nucleic acid enzyme selection*. Biol Chem, 2001. **382**(9): p. 1327-34.

21. Gschwind, A., O.M. Fischer, and A. Ullrich, *The discovery of receptor tyrosine kinases: targets for cancer therapy*. Nat Rev Cancer, 2004. **4**(5): p. 361-70.

22. Hansford, J.R. and L.M. Mulligan, *Multiple endocrine neoplasia type 2 and RET: from neoplasia to neurogenesis*. J Med Genet, 2000. **37**(11): p. 817-27.

23. Ichihara, M., Y. Murakumo, and M. Takahashi, *RET and neuroendocrine tumors*. Cancer Lett, 2004. **204**(2): p. 197-211.

24. Cerchia, L., et al., *Neutralizing aptamers from whole-cell SELEX inhibit the RET receptor tyrosine kinase*. PLoS Biol, 2005. **3**(4): p. e123.

25. Petitclerc, E., et al., *Integrin alpha(v)beta3 promotes M21 melanoma growth in human skin by regulating tumor cell survival*. Cancer Res, 1999. **59**(11): p. 2724-30.

26. Mi, J., et al., *Targeted inhibition of alphavbeta3 integrin with an RNA aptamer impairs endothelial cell growth and survival*. Biochem Biophys Res Commun, 2005. **338**(2): p. 956-63.

27. Adai, A.T., et al., *LGL: creating a map of protein function with an algorithm for visualizing very large biological networks*. J Mol Biol, 2004. **340**(1): p. 179-90.

28. Almeida, J.S., et al., *Analysis of genomic sequences by Chaos Game Representation*. Bioinformatics, 2001. **17**(5): p. 429-37.

29. Brazma, A., et al., *Minimum information about a microarray experiment (MIAME)-toward standards for microarray data*. Nat Genet, 2001. **29**(4): p. 365-71.

30. Brody, E.N. and L. Gold, *Aptamers as therapeutic and diagnostic agents*. J Biotechnol, 2000. **74**(1): p. 5-13.

31. Burgstaller, P., A. Jenne, and M. Blind, *Aptamers and aptazymes: accelerating small molecule drug discovery*. Curr Opin Drug Discov Devel, 2002. **5**(5): p. 690-700.

32. Chen, N., et al., *WormBase: a comprehensive data resource for Caenorhabditis biology and genomics*. Nucleic Acids Res, 2005. **33**(Database issue): p. D383-9.

33. Ding, Y. and C.E. Lawrence, *Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond*. Nucleic Acids Res, 2001. **29**(5): p. 1034-46.

34. Eaton, B.E., *The joys of in vitro selection: chemically dressing oligonucleotides to satiate protein targets*. Curr Opin Chem Biol, 1997. **1**(1): p. 10-6.

35. Ekland, E.H., J.W. Szostak, and D.P. Bartel, *Structurally complex and highly active RNA ligases derived from random RNA sequences*. Science, 1995. **269**(5222): p. 364-70.

36. Famulok, M. and G. Mayer, *Aptamers as tools in molecular biology and immunology*. Curr Top Microbiol Immunol, 1999. **243**: p. 123-36.

37. Feigon, J., T. Dieckmann, and F.W. Smith, *Aptamer structures from A to zeta*. Chem Biol, 1996. **3**(8): p. 611-7.

38. Galperin, M.Y., *The Molecular Biology Database Collection: 2005 update*. Nucleic Acids Res, 2005. **33**(Database issue): p. D5-24.

39. Gelbart, W.M., et al., *FlyBase: a Drosophila database. The FlyBase consortium*. Nucleic Acids Res, 1997. **25**(1): p. 63-6.

40. Harris, M.A., et al., *The Gene Ontology (GO) database and informatics resource*. Nucleic Acids Res, 2004. **32**(Database issue): p. D258-61.

41. Hofacker, I., et al., *Fast Folding and Comparison of RNA Secondary Structures*. Monatsh. Chem., 1994. **125**: p. 167-188.

42. Hokamp, K. and K. Wolfe, *What's new in the library? What's new in GenBank? let PubCrawler tell you*. Trends Genet, 1999. **15**(11): p. 471-2.

43. Huang, F., Z. Yang, and M. Yarus, *Self-capping RNA catalysts derived from selection-amplification*. Biol Bull, 1999. **196**(3): p. 320-1.

44. Hughes, J.D., et al., *Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae*. J Mol Biol, 2000. **296**(5): p. 1205-14.

45. Illangasekare, M. and M. Yarus, *A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis*. Rna, 1999. **5**(11): p. 1482-9.

46. Jayasena, S.D., *Aptamers: an emerging class of molecules that rival antibodies in diagnostics*. Clin Chem, 1999. **45**(9): p. 1628-50.

47. Jeffrey, H.J., *Chaos game representation of gene structure*. Nucleic Acids Res, 1990. **18**(8): p. 2163-70.

48. Kanehisa, M., et al., *The KEGG resource for deciphering the genome*. Nucleic Acids Res, 2004. **32**(Database issue): p. D277-80.

49. Klein, R.J., Z. Misulovin, and S.R. Eddy, *Noncoding RNA genes identified in AT-rich hyperthermophiles*. Proc Natl Acad Sci U S A, 2002. **99**(11): p. 7542-7.

50. Kunsch, C., S.M. Ruben, and C.A. Rosen, *Selection of optimal kappa B/Rel DNA-binding motifs: interaction of both subunits of NF-kappa B with DNA is required for transcriptional activation*. Mol Cell Biol, 1992. **12**(10): p. 4412-21.

51. Lebruska, L.L. and L.J. Maher, 3rd, *Selection and characterization of an RNA decoy for transcription factor NF-kappa B*. Biochemistry, 1999. **38**(10): p. 3168-74.

52. Lee, J.F., et al., *Aptamer database*. Nucleic Acids Res, 2004. **32**(Database issue): p. D95-100.

53. Li, W., B. Ma, and B. Shapiro, *Molecular dynamics simulations of the denaturation and refolding of an RNA tetraloop*. J Biomol Struct Dyn, 2001. **19**(3): p. 381-96.

54. Lupold, S.E., et al., *Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen*. Cancer Res, 2002. **62**(14): p. 4029-33.

55. Martell, R.E., J.R. Nevins, and B.A. Sullenger, *Optimizing aptamer activity for gene therapy applications using expression cassette SELEX*. Mol Ther, 2002. **6**(1): p. 30-4.

56. McCaskill, J.S., *The equilibrium partition function and base pair binding probabilities for RNA secondary structure*. Biopolymers, 1990. **29**(6-7): p. 1105-19.

57. Meyers, L.A., et al., *The robustness of naturally and artificially selected nucleic acid secondary structures*. J Mol Evol, 2004. **58**(6): p. 681-91.

58. Morgan, W.D., et al., *RNA sequence and secondary structure requirements for rho-dependent transcription termination*. Nucleic Acids Res, 1985. **13**(10): p. 3739-54.

59. Morgan, W.D., D.G. Bear, and P.H. von Hippel, *Rho-dependent termination of transcription. I. Identification and characterization of termination sites for transcription from the bacteriophage lambda PR promoter*. J Biol Chem, 1983. **258**(15): p. 9553-64.

60. Parkinson, H., et al., *ArrayExpress--a public repository for microarray gene expression data at the EBI*. Nucleic Acids Res, 2005. **33**(Database issue): p. D553-5.

61. Ponomarenko, J.V., et al., *SELEX_DB: a database on in vitro selected oligomers adapted for recognizing natural sites and for analyzing both SNPs and site-directed mutagenesis data*. Nucleic Acids Res, 2002. **30**(1): p. 195-9.

62. Rhee, S.Y., et al., *The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community*. Nucleic Acids Res, 2003. **31**(1): p. 224-8.

63. Roulet, E., et al., *High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites*. Nat Biotechnol, 2002. **20**(8): p. 831-5.

64. Sorin, E.J., et al., *RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop*. J Mol Biol, 2002. **317**(4): p. 493-506.

65. Sudarsanam, P., Y. Pilpel, and G.M. Church, *Genome-wide co-occurrence of promoter elements reveals a cis-regulatory cassette of rRNA transcription motifs in Saccharomyces cerevisiae*. Genome Res, 2002. **12**(11): p. 1723-31.

66. Unrau, P.J. and D.P. Bartel, *RNA-catalysed nucleotide synthesis*. Nature, 1998. **395**(6699): p. 260-3.

67. Varani, G., *Exceptionally stable nucleic acid hairpins*. Annu Rev Biophys Biomol Struct, 1995. **24**: p. 379-404.

68. Williams, D.J. and K.B. Hall, *Experimental and computational studies of the G[UUCG]C RNA tetraloop*. J Mol Biol, 2000. **297**(5): p. 1045-61.

69. Winkler, W., A. Nahvi, and R.R. Breaker, *Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression*. Nature, 2002. **419**(6910): p. 952-6.

70. Winkler, W.C., S. Cohen-Chalamish, and R.R. Breaker, *An mRNA structure that controls gene expression by binding FMN*. Proc Natl Acad Sci U S A, 2002. **99**(25): p. 15908-13.

71. Wright, W.E. and W.D. Funk, *CASTing for multicomponent DNA-binding complexes*. Trends Biochem Sci, 1993. **18**(3): p. 77-80.

72. Zuker, M. and P. Stiegler, *Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information*. Nucleic Acids Res, 1981. **9**(1): p. 133-48.

73. Bakhanashvili, M. and A. Hizi, *Fidelity of the RNA-dependent DNA synthesis exhibited by the reverse transcriptases of human immunodeficiency virus types 1 and 2 and of murine leukemia virus: mispair extension frequencies*. Biochemistry, 1992. **31**(39): p. 9393-8.

74. Battula, N. and L.A. Loeb, *On the fidelity of DNA replication. Characterization of polynucleotides with errors in base-pairing synthesized by avian myeloblastosis virus deoxyribonucleic acid polymerase*. J Biol Chem, 1975. **250**(12): p. 4405-9.

75. Benkestock, K., et al., *Automated nano-electrospray mass spectrometry for protein-ligand screening by noncovalent interaction applied to human H-FABP and A-FABP*. J Biomol Screen, 2003. **8**(3): p. 247-56.

76. Boland, E.J., et al., *Automation of the Maxam-Gilbert chemical sequencing reactions*. Biotechniques, 1994. **16**(6): p. 1088-92, 1094-5.

77. Chapman, T., *Lab automation and robotics: Automation on the move*. Nature, 2003. **421**(6923): p. 661, 663, 665-6.

78. Chen, B.Y., H.W. Janes, and S. Chen, *Computer programs for PCR primer design and analysis*. Methods Mol Biol, 2002. **192**: p. 19-29.

79. Civitello, A., Richards, S., Gibbs, RA., *A simple protocol for the automation of DNA cycle sequencing reactions and polymerase chain reactions*. DNA Seq., 1992. **3**(1): p. 17-23.

80. Clark, S.L. and V.T. Remcho, *Electrochromatographic retention studies on a flavin-binding RNA aptamer sorbent*. Anal Chem, 2003. **75**(21): p. 5692-6.

81. Cowan, J.A., et al., *Recognition of a cognate RNA aptamer by neomycin B: quantitative evaluation of hydrogen bonding and electrostatic interactions*. Nucleic Acids Res, 2000. **28**(15): p. 2935-42.

82. Cox, J.C., et al., *Automated acquisition of aptamer sequences*. Comb Chem High Throughput Screen, 2002. **5**(4): p. 289-99.

83. Cox, J.C., P. Rudolph, and A.D. Ellington, *Automated RNA selection*. Biotechnol Prog, 1998. **14**(6): p. 845-50.

84. Crameri, A. and W.P. Stemmer, *10(20)-fold aptamer library amplification without gel purification*. Nucleic Acids Res, 1993. **21**(18): p. 4410.

85. Dapic, V., et al., *Biophysical and biological properties of quadruplex oligodeoxyribonucleotides*. Nucleic Acids Res, 2003. **31**(8): p. 2097-107.

86. Draper, D.E., *A guide to ions and RNA structure*. Rna, 2004. **10**(3): p. 335-43.

87. Ellington, A.D., *Purification of Oligonucleotides Using Denaturing Polyacrylamide Gel Electrophoresis*, in *Short protocols in molecular biology*., e.a. F. Ausubel, Editor. 1997, John Wiley & Sons, Inc.: New York.

88. Felsenstein, J., *An alternating least squares approach to inferring phylogenies from pairwise distances*. Syst Biol, 1997. **46**(1): p. 101-11.

89. Grundy, F.J. and T.M. Henkin, *From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements*. Crit Rev Biochem Mol Biol, 2006. **41**(6): p. 329-38.

90. Harada, K. and L.E. Orgel, *In vitro selection of optimal DNA substrates for T4 RNA ligase*. Proc Natl Acad Sci U S A, 1993. **90**(4): p. 1576-9.

91. Hesselberth, J., et al., *In vitro selection of nucleic acids for diagnostic applications*. J Biotechnol, 2000. **74**(1): p. 15-25.

92. Hesselberth, J.R., et al., *In vitro selection of RNA molecules that inhibit the activity of ricin A-chain*. J Biol Chem, 2000. **275**(7): p. 4937-42.

93. Jhaveri, S.a.E.A., *In vitro selection of RNA aptamers to a protein target by filter immobilization*, in *Current Protocols in Nucleic Acid Chemistry*, e.a. S.L. Beaucage, Editor. 2000, John Wiley and Sons, Inc.: New York. p. 9.3.1-9.3.25.

94. King, R.D., et al., *Functional genomic hypothesis generation and experimentation by a robot scientist*. Nature, 2004. **427**(6971): p. 247-52.

95. Lee, J.F., Cox, J.C., Collett, JR., Ellington, AD, *Exploring Sequence Space Through Automated Aptamer Selections*. Journal of the Association for Laboratory Automation, 2005. **10**(4): p. 213-218.

96. Malmgren, C., et al., *An antisense/target RNA duplex or a strong intramolecular RNA structure 5' of a translation initiation signal blocks ribosome binding: the case of plasmid R1*. Rna, 1996. **2**(10): p. 1022-32.

97. McPherson, J.D., et al., *A physical map of the human genome*. Nature, 2001. **409**(6822): p. 934-41.

98. Mendelman, L.V., J. Petruska, and M.F. Goodman, *Base mispair extension kinetics. Comparison of DNA polymerase alpha and reverse transcriptase*. J Biol Chem, 1990. **265**(4): p. 2338-46.

99. Parrott, A.M., et al., *RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour*. Nucleic Acids Res, 2000. **28**(2): p. 489-97.

100. Pollard J., S.D.B., and A.D. Ellington, *Design, synthesis, and amplification of DNA pools for in vitro selection*, in *Current Protocols in Nucleic Acid Chemistry*, e.a. S.L. Beaucage, Editor. 2000, John Wiley & Sons: New York. p. 9.2.1-9.2.23.

101. Schmitt, T. and N. Lehman, *Non-unity molecular heritability demonstrated by continuous evolution in vitro*. Chem Biol, 1999. **6**(12): p. 857-69.

102. Venter, J.C., et al., *The sequence of the human genome*. Science, 2001. **291**(5507): p. 1304-51.

103. Zimmermann, G.R., et al., *Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer*. Rna, 2000. **6**(5): p. 659-67.

104. Brookfield, J., *Fitness*. Curr Biol, 2002. **12**(21): p. R717.

105. Wright, S., *General, Group and Special Size Factors*. Genetics, 1932. **17**(5): p. 603-19.

106. Fischer KH, H.J., *Spin glass*. 1991, Cambridge: Cambridge University Press.

107. Kauffman, S.A., Perelson, AS, *Molecular evolution on rugged landscapes: Santa Fe Institute studies in the sciences of complexity*. 1991.

108. Fontana, W. and P. Schuster, *Continuity in evolution: on the nature of transitions*. Science, 1998. **280**(5368): p. 1451-5.

109. Salehi-Ashtiani, K. and J.W. Szostak, *In vitro evolution suggests multiple origins for the hammerhead ribozyme*. Nature, 2001. **414**(6859): p. 82-4.

110. Ellington, A.D. and J.W. Szostak, *Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures*. Nature, 1992. **355**(6363): p. 850-2.

111. Hamming, R.W., *Coding and Information Theory*. 2nd ed. 1986: Prentice Hall.

112. Kauffman, S.A. and E.D. Weinberger, *The NK model of rugged fitness landscapes and its application to maturation of the immune response*. J Theor Biol, 1989. **141**(2): p. 211-45.

113. Higgins, D.G., *Sequence ordinations: a multivariate analysis approach to analysing large sequence data sets*. Comput Appl Biosci, 1992. **8**(1): p. 15-22.

114. Clakson, T., D. D. Gussow, and P. T. Jones, *Characterizing recombinant clones directly by 'PCR screening'*, in *PCR 1*, D.R.a.B.D. Hames, Editor. 1991, Oxford University Press: Oxford. p. 210-212.

115. Learn, G.H., Jr., et al., *Maintaining the integrity of human immunodeficiency virus sequence databases*. J Virol, 1996. **70**(8): p. 5720-30.

116. Joliffe, I.T., *Principal Component Analysis*. 1986, New York.

117. Wold, S., *Estimation of principal component and related models by iterative least squares*. Multivariate Analysis, ed. P. Krishnaiah. 1966, New York: Academic Press.

118. Brambilla, E., et al., *The new World Health Organization classification of lung tumours*. Eur Respir J, 2001. **18**(6): p. 1059-68.

119. Bharti, A., P.C. Ma, and R. Salgia, *Biomarker discovery in lung cancer-promises and challenges of clinical proteomics*. Mass Spectrom Rev, 2007.

120. Chute, J.P., et al., *Twenty years of phase III trials for patients with extensive-stage small-cell lung cancer: perceptible progress*. J Clin Oncol, 1999. **17**(6): p. 1794-801.

121. Gail, M.H., et al., *Multiple markers for lung cancer diagnosis: validation of models for localized lung cancer*. J Natl Cancer Inst, 1988. **80**(2): p. 97-101.

122. Sander, C., *Genomic medicine and the future of health care*. Science, 2000. **287**(5460): p. 1977-8.

123. Brundage, M.D., D. Davies, and W.J. Mackillop, *Prognostic factors in non-small cell lung cancer: a decade of progress*. Chest, 2002. **122**(3): p. 1037-57.

124. McCarthy, N.J. and S.M. Swain, *Tumor markers: should we or shouldn't we?* Cancer J, 2001. **7**(3): p. 175-7.

125. Strauss, G.M. and A.T. Skarin, *Use of tumor markers in lung cancer*. Hematol Oncol Clin North Am, 1994. **8**(3): p. 507-32.

126. Ma, P.C., et al., *Circulating tumor cells and serum tumor biomarkers in small cell lung cancer*. Anticancer Res, 2003. **23**(1A): p. 49-62.

127. Shapiro, G.I., et al., *Reciprocal Rb inactivation and p16INK4 expression in primary lung cancers and cell lines*. Cancer Res, 1995. **55**(3): p. 505-9.

128. Zereu, M., J.J. Vinholes, and C.G. Zettler, *p53 and Bcl-2 protein expression and its relationship with prognosis in small-cell lung cancer*. Clin Lung Cancer, 2003. **4**(5): p. 298-302.

129. Dwarakanath, S., et al., *Quantum dot-antibody and aptamer conjugates shift fluorescence upon binding bacteria*. Biochem Biophys Res Commun, 2004. **325**(3): p. 739-43.

130. Farokhzad, O.C., J.M. Karp, and R. Langer, *Nanoparticle-aptamer bioconjugates for cancer targeting*. Expert Opin Drug Deliv, 2006. **3**(3): p. 311-24.

131. Famulok, M. and G. Mayer, *Chemical biology: aptamers in nanoland*. Nature, 2006. **439**(7077): p. 666-9.

132. Cerchia, L., et al., *Nucleic acid aptamers in cancer medicine*. FEBS Lett, 2002. **528**(1-3): p. 12-6.

133. Blank, M., et al., *Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels. selective targeting of endothelial regulatory protein pigpen*. J Biol Chem, 2001. **276**(19): p. 16464-8.

134. Daniels, D.A., et al., *A tenascin-C aptamer identified by tumor cell SELEX: systematic evolution of ligands by exponential enrichment*. Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15416-21.

135. Hicke, B.J., et al., *Tenascin-C aptamers are generated using tumor cells and purified protein*. J Biol Chem, 2001. **276**(52): p. 48644-54.

136. Chu, T.C., et al., *Aptamer:toxin conjugates that specifically target prostate tumor cells*. Cancer Res, 2006. **66**(12): p. 5989-92.

137. Chu, T.C., et al., *Labeling tumor cells with fluorescent nanocrystal-aptamer bioconjugates*. Biosens Bioelectron, 2006. **21**(10): p. 1859-66.

138. Alliegro, M.C. and M.A. Alliegro, *Nuclear injection of anti-pigpen antibodies inhibits endothelial cell division*. J Biol Chem, 2002. **277**(21): p. 19037-41.

139. Shangguan, D., et al., *Aptamers evolved from live cells as effective molecular probes for cancer study*. Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11838-43.

140. Fukuoka, M., et al., *Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]*. J Clin Oncol, 2003. **21**(12): p. 2237-46.

141. Kris, M.G., et al., *Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial*. Jama, 2003. **290**(16): p. 2149-58.

142. Shepherd, F.A., et al., *Erlotinib in previously treated non-small-cell lung cancer*. N Engl J Med, 2005. **353**(2): p. 123-32.

143. Malkinson, A.M., et al., *Mouse lung epithelial cell lines--tools for the study of differentiation and the neoplastic phenotype*. Toxicology, 1997. **123**(1-2): p. 53-100.

144. Homann, M. and H.U. Goringer, *Uptake and intracellular transport of RNA aptamers in African trypanosomes suggest therapeutic "piggy-back" approach*. Bioorg Med Chem, 2001. **9**(10): p. 2571-80.

145. Lorger, M., et al., *Targeting the variable surface of African trypanosomes with variant surface glycoprotein-specific, serum-stable RNA aptamers*. Eukaryot Cell, 2003. **2**(1): p. 84-94.

146. Morris, K.N., et al., *High affinity ligands from in vitro selection: complex targets*. Proc Natl Acad Sci U S A, 1998. **95**(6): p. 2902-7.

147. Hicke, L., *A new ticket for entry into budding vesicles-ubiquitin*. Cell, 2001. **106**(5): p. 527-30.

Vita

Jennifer Fang-En Lee was born in Taipei, Taiwan on August 23, 1977, the daughter of Rung-Chang Lee and Shiaw-Feng Hsu-Lee. At the age of five, she moved to Argentina with her parents and her two brothers, David and Alex, where she lived for ten years. She moved to the United States in July of 1992 and attended Cypress Creek High School in Orlando, Florida. After high school in 1996 she attended the University of Florida where she received her Bachelor of Science in Biochemistry with Highest Honors in 1999. She began her research career as an undergraduate student by joining the Laboratory of Profesors Maureen Goodenow and John Sleasman by conducting research on infectious diseases. After graduation she continued working as a laboratory research. In August of 2000, she entered the graduate program in the Department of Cell and Molecular Biology at the University of Texas, Austin and started her Ph.D. program under the tutelage and supervision of Professor Andrew D. Ellington. Her scientific publications include:

Publications

- **Lee, JF**, Hesselberth JR, Meyers LA, Ellington AD.
  Aptamer database. Nucleic Acids Res. 2004 Jan 1;32 Database issue: D95100.
- Meyers LA**, Lee JF**, Cowperthwaite M, Ellington AD.
  The robustness of naturally and artificially selected nucleic acid secondary structures. J Mol Evol. 2004 Jun;58(6):68191.
- Collett JR, Cho EJ**, Lee JF,** Levy M, Hood AJ, Wan C, Ellington AD.
  Functional RNA microarrays for highthroughput screening of antiprotein aptamers. Anal Biochem. 2005 Mar 1;338(1):11323.

- **Lee, JF.,** Cox, JC., Collett, JR., Ellington, AD.
  Exploring Sequence Space Through Automated Aptamer Selection
  Journal of the Association for Laboratory Automation 2005. 10(4):213218.

- **Lee JF**, Stovall GM, Ellington AD.
  Aptamer therapeutics advance. (review). Curr Opin Chem Biol. 2006 Jun;10(3):2829.

Book Chapter

- Ellington, AD, Cox JC, **Lee JF**, Collett JR.
  Automated in vitro selection and microarray applications for functional RNA sequences, in The RNA World, 3rd Ed. 2005, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York. p. 683719.

Permanent address: 3524 Greystone Dr. #163 Austin TX 78731

This dissertation was typed by Jennifer F. Lee.